Cross-sectional forecasts of the equity premium

Samuel Brodsky Thompson
Harvard University
(joint with Christopher Polk and Tuomo Vuolteenaho)

Presentation at Research Workshop in Applied Statistics
10/20/2004

Can we predict the equity premium?

- The equity premium is the return on stocks minus the return on bonds
- We provide a new theoretically-motivated class of variables that can forecast the equity premium
- We provide a new statistical method for small-sample inference in predictive regressions
- We provide new empirical evidence that market returns are predictable in the US as well as in other countries

Stock market predictability?

- Perhaps the price/earnings ratio (pe) predicts near-term stock returns.
- Suppose the pe ratio is a stationary random variable. If it is above its mean there are two ways for it to mean-revert:
  - Price falls, so returns are bad.
  - Earnings rise, so the optimists were right.
- The historical record suggests that earnings are unpredictable, so on average high pe's lead to low returns.
- Simple regression tests suggests that the pe ratio predicts future stock returns.

Cross sectional versus time series predictors

- The popular predictors, like the dividend yield and price-earnings ratio, ignore a lot of information in the cross-section of stock prices. How do we use that information?
- We propose a cross-sectional predictor of market returns. Using accounting variables, we determine which stocks are currently “good buys” relative to others. If those stocks have recently been correlated with market movements, we expect the market to do well. If the “bad buys” are relatively more correlated with market movements, we expect the market to do poorly.

CAPM links the time series to cross-section

The CAPM:

$$\left(R_{it} - R_{rf,t}\right) = \beta_i \left(R_{Mt} - R_{rf,t}\right) + e_{it}$$

R_{it} = return in month t on stock i

$R_{rf,t}$ = return on risk-free asset (T-bill)

R_{Mt} = return on entire stock market

R_{it}- $R_{rf,t}$ is the equity premium

Stock returns are decomposed into a common market-wide component R_{Mt}-$R_{rf,t}$ and an idiosyncratic component e_{it}.

Our predictor backs out the price of risk from the cross section

The CAPM:

$$E_{t-1}\left(R_{it} - R_{rf,t}\right) = \beta_i E_{t-1}\left(R_{Mt} - R_{rf,t}\right)$$

1. Start with cross-sectional forecasts of individual stock returns, $E_{t-1}R_{it}$.
2. Construct estimates of β_i.
3. Regress $E_{t-1}R_{it}$ on β_i. The slope is an estimate of $\lambda_t = E_{t-1}R_{Mt} - R_{rf}$, the market equity premium.

CAPM interpretation of the predictor

- When high β stocks have high expected returns, the market has high expected returns.
- According to the CAPM, investors want to be compensated for bearing risks correlated with the market. If the stocks that are highly correlated with the market are cheap, then investors are scared of some market-wide risk factor, and demand a higher expected return from the market.
- The CAPM is a model of insurance. Low β stocks do well when the market falls, so offer insurance. The insurance feature increases the price and decreases the return.

Constructing the predictor

- Our predictor is the slope coefficient from regressing $E_{t-1}R_{it}$ on β_i.
- β_i is estimated from 3 year rolling regressions.
- $E_{t-1}R_{it}$ is estimated in several ways:
  - Regress past returns for individual stocks on predictors like the earnings yield, then use the fitted value as a predictor
  - Combine accounting information with the Gordon growth model and Campbell-Shiller decompositions:

    $E_{t-1}R_{it}$=(dividend yield)$_{it-1}$ + (growth proxies)$_{it-1}$
  - Subtract beta for growth portfolio from beta on value portfolio.

Statistical problem

A statistical problem

- We estimate the following regressions:

$$R^e_{M,t} = \mu_1 + \underbrace{.0908}_{\text{t-stat } 3.605} \lambda^{REG}_{t-1} + u_t$$

$$R^e_{M,t} = \mu_1 + \underbrace{.0170}_{\text{t-stat } 3.454} ep_{t-1} + u_t$$

$$R^e_{M,t} = \mu_1 + \underbrace{.066}_{\text{t-stat } 2.36} \lambda^{REG}_{t-1} + \underbrace{.012}_{\text{t-stat } 2.12} ep_{t-1} + u_t$$

- Everything is significant using the usual asymptotic critical values – but we know that classical asymptotics offer a poor approximation to the null distribution of these test statistics.
- This leads to an interesting statistical problem. How do we get good p-values?

More structure

$$y_t = \mu_1 + \theta x_{t-1} + u_t$$

$$x_t = \mu_2 + \rho x_{t-1} + v_t$$

$$Eu_t = Ev_t = 0, \text{corr}(u_t, v_t) = \gamma, Eu_t^2 = \sigma_u^2, Ev_t^2 = \sigma_v^2$$

- y_t is the monthly excess return (over T-bill) on the stock market
- x_t is a predictor variable such as log(earnings/price), so we expect that high x_t leads to high returns.
- We wish to test the null hypothesis H: $\theta=0$ versus the alternative K: $\theta>0$.

The t-test is biased!

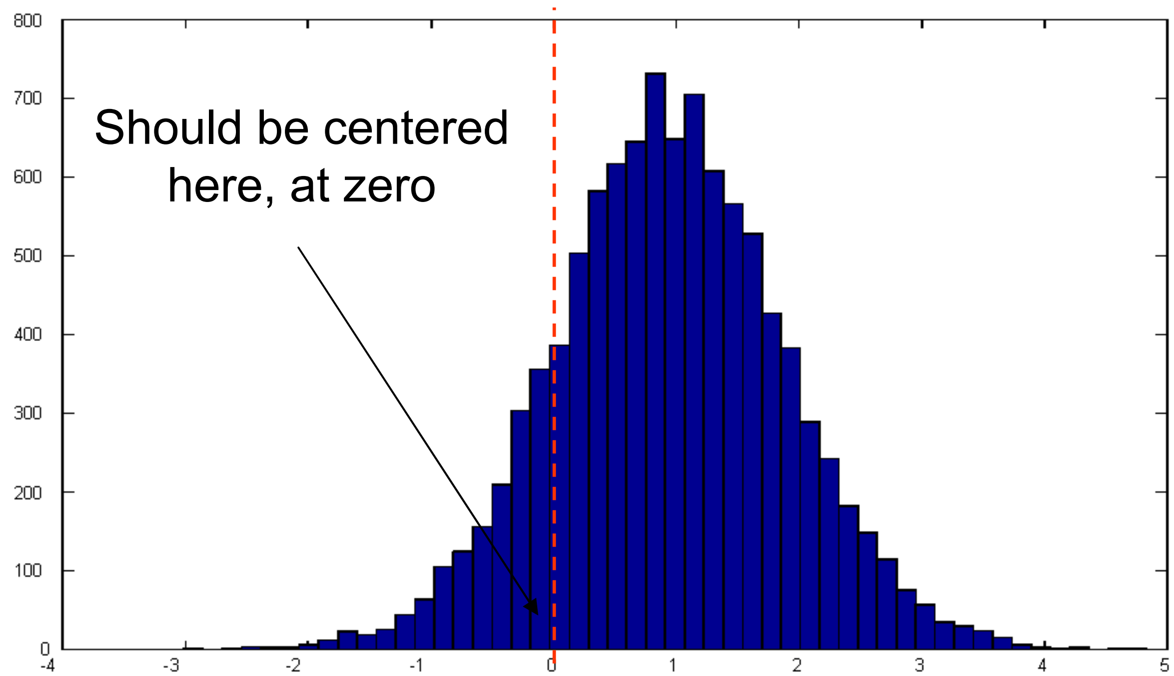
- We reject the null θ=0 for values of the t-statistic

$$\hat{t} = \hat{\sigma}_u^{-1}\hat{\theta}\sqrt{\sum (x_{t-1} - \overline{x})^2}$$

  that are higher than the critical value
- Classical asymptotic theory states that in a large sample the t-statistic is approximately N(0,1) under the null. But this is a terrible approximation
- The null distribution of the t-statistic is centered at a positive number, leading to over-rejection of a true null hypothesis
- So, asymptotics can indicate predictability even when there is none

A Monte Carlo

- Histogram of OLS estimates calibrated to Stambaugh's estimates using the dividend yield: ρ=.972, γ=-.90, T=120.
- The simulation is carried out under the null, so θ=0. If classical asymptotics are working, the estimator should be centered at zero.



Bias can change the answer

Stambaugh : Predicting with the dividend yield, $\rho = 0.972$ and $\gamma = -.90$

	Sample period			
	1927–1996	1927–1951	1952–1996	1977–1996
A. True properties				
Bias	0.07	0.18	0.18	0.42
Standard deviation	0.16	0.33	0.27	0.45
Skewness	0.71	0.83	0.98	1.29
Kurtosis	3.84	4.14	4.62	5.83
p-value for $\beta = 0$	0.17	0.42	0.15	0.64
B. Properties in the standard regression setting				
Bias	0	0	0	0
Standard deviation	0.14	0.27	0.20	0.30
Skewness	0	0	0	0
Kurtosis	3	3	3	3
p-value for $\beta = 0$	0.06	0.22	0.02	0.26

When does normality fail?

$$y_t = \mu_1 + \theta x_{t-1} + u_t$$

$$x_t = \mu_2 + \rho x_{t-1} + v_t$$

$$Eu_t = Ev_t = 0, \operatorname{corr}(u_t, v_t) = \gamma, Eu_t^2 = \sigma_u^2, Ev_t^2 = \sigma_v^2$$

- The ols estimate and t-stat for θ are biased when ρ is close to one and Corr(u,v) is nonzero.
- This is exactly what happens when we predict stock returns with earnings/price. earnings/price is highly autocorrelated, so ρ is big. When returns are good, u is positive and price rises. Since earnings don't move much over short periods, earnings/price falls, inducing a negative correlation between u and v.

Intuition by Jeremy Siegel

- Examine a chart of the price of a stock over the past year. Then record all past prices, rank these prices from highest to lowest, and measure the subsequent return on the stock following each price observation.
- It is easy to see that returns will be poor following high stock prices and the return will be high following low stock prices.
- But this is because we have, after the fact, identified the high and low prices for the stock. Once the high price has been identified, all subsequent prices must be lower, assuring poor returns.

Intuition by Jeremy Siegel

- In 1993 the yields on stock indexes approached historic lows.
- But if yields go even lower, then the year 1993 will no longer stand out as a year of extremely low yields. (This happened!)
- When the data are stationary and not persistent, we can figure out what the "average" or "typical" yield should be. But it the yield is persistent, it takes a huge amount of data to figure out which dividend yields are low by historical standards.

Some statistical intuition

$$T^{-1}\sum\left[(y_t - \bar{y}_T)(x_{t-1} - \bar{x}_T)\right]$$
$$= T^{-1}\sum\left[(y_t x_{t-1}) - (\bar{y}_T \bar{x}_T)\right]$$

In small samples (size T),

$$E(\bar{y}_T \bar{x}_T) < E(\bar{y}_T)E(\bar{x}_T) = E(y)E(x)$$

if $\gamma < 0$ and $\rho > 0$

This is an example of a nuisance parameter problem

Local to unity asymptotics:

$\rho = 1 + c/T$ as $T \to \infty$.

$dW_c(s) = cW_c(s)ds + dW_c$

$$tstat \to \gamma \frac{\int \left(W_c - \int W_c\right) dW_0}{\sqrt{\int \left(W_c - \int W_c\right)^2}} + \sqrt{1-\gamma^2} N(0,1)$$

SOLUTION

Conditional p-values

$$y_t = \mu_1 + \theta x_{t-1} + u_t$$
$$x_t = \mu_2 + \rho x_{t-1} + v_t$$

\+ homoskedastic normal errors

- The distribution of the OLS estimator for θ is biased (upward, usually).
- The distribution depends on ρ, a nuisance parameter that we estimate poorly. What can we do?
- Jansson and Moreira (2002) have a solution – conditional inference:
  - Find a sufficient statistics for ρ.
  - The conditional null distribution, given the sufficient statistics, does not depend on ρ.
  - So test using the conditional null distribution.

Simplified example

Suppose $\sigma_e^2 = 1$, $\sigma_u^2 = 1$ and γ is known.

Suppose there are no intercepts. The likelihood function is

$$p(\theta,\rho) = K\exp\left[-\frac{1}{2(1-\gamma^2)}\left\{\sum(y_t-\theta x_{t-1})^2+\sum(x_t-\rho x_{t-1})^2-2\gamma\sum(y_t-\theta x_{t-1})(x_t-\rho x_{t-1})\right\}\right]$$

- Step 1: impose the null that $\theta = 0$.

$$p(0,\rho) = K\exp\left[-\frac{1}{2(1-\gamma^2)}\left\{\sum y_t^2+\sum(x_t-\rho x_{t-1})^2-2\gamma\sum y_t(x_t-\rho x_{t-1})\right\}\right]$$

- Step 2: By the factorization theorem,

$S = \left(\sum x_{t-1}^2, \sum x_{t-1}(x_t-\gamma y_t)\right)$ are sufficient statistics for ρ:

$$p(0,\rho) = K\exp\left[-\frac{1}{2(1-\gamma^2)}\left\{\sum y_t^2+\sum x_t^2-2\gamma\sum y_t x_t+\rho^2\sum x_{t-1}^2-2\rho\sum x_{t-1}(x_t-\gamma y_t)\right\}\right]$$

$$p(0,\rho) = Kg(S,\rho)h(x_1,\ldots,x_T,y_1,\ldots,y_T)$$

- Step 3: Recall the definition of a sufficient statistic:

the conditional distribution of anything given S does not depend on ρ.

So reject the null when $\hat{t}$ is bigger than the 95% quantile of the null distribution, conditional on S.

This test will have the correct null rejection probability for any sample size and for any value of ρ.

Conditional quantiles

- We want q(s,α), the α-quantile of the conditional null of the t-stat given S=s:

$$\text{probability}\left[\hat{t} > q(s,\alpha) \middle| S = s, \theta = 0\right] = \alpha$$

- If we reject when the t-stat is greater than q(S,.95), we reject a true null 5% of the time for any ρ and n.
- There is no closed-form solution for the quantile.
- We've got to *nonparametrically* estimate an *extreme quantile* of a *conditional* distribution.

    In the words of Ariel Pakes:

    "Every part of that problem is hard."

Local nonparametrics

- Local nonparametric estimation.
  - Simulate a lot of data sets
  - Throw out all the data where the realized sufficient statistic are far from the observed values
  - Sort the simulated t-statistics.
  - The 95% conditional quantile estimate is the 95% empirical quantile of the sorted t-statistics.
- Pros:
  - Statistical theorems show that it is consistent.
- Cons:
  - How do you Monte Carlo it? It will take forever.
  - To construct a confidence interval you've got to do it for each null value.

We tried it and it is really slow.

Global nonparametrics

- Consider a global approximation to $q(s,\alpha)$. For example, we could approximate $q(s,\alpha)$ with a polynomial in s.
- Local approximations are straightforward to estimate. But they are hard to communicate – you have got to give a researcher all your simulated draws. They are also slow.
- Global approximations are hard to estimate. How do we choose the coefficients of the polynomial? But they are easy to communicate. Once we've got the coefficients of our polynomial approximation, anyone can use it.
- If we want to get applied people to carry out conditional inference, we should provide a global approximation to the conditional distribution.

Approximating q(S,α) with a neural network

$$q(S,\alpha) \approx q_\alpha^{nn}(X,\hat{\psi},\hat{\xi})$$

$$q_\alpha^{nn}(X,\psi,\xi) \equiv \text{sign}(\hat{\gamma})\mu(X) + \sigma(X)\Phi^{-1}(\alpha)$$

$$\mu(X) = \xi_0^\mu + \sum_{j=1}^4 \xi_j^\mu g(\psi_j' e^X)$$

$$\sigma(X) = \exp\left[\xi_0^\sigma + \sum_{j=1}^4 \xi_j^\sigma g(\psi_j' e^X)\right]$$

$$X \equiv \begin{bmatrix} 0 & \dfrac{T(\hat{\rho}_R - 1)}{50} & -\dfrac{\sum (x_{t-1} - \overline{x})^2}{T^2 \hat{\sigma}_v^2} & \log|\hat{\gamma}| & -\dfrac{T}{100} \end{bmatrix}$$

$$\hat{\rho}_R \equiv \sum \left[(x_{t-1} - \overline{x})(x_t - \hat{\gamma}\tfrac{\hat{\sigma}_v}{\hat{\sigma}_u} y_t)\right] \Big/ \sum (x_{t-1} - \overline{x})^2$$

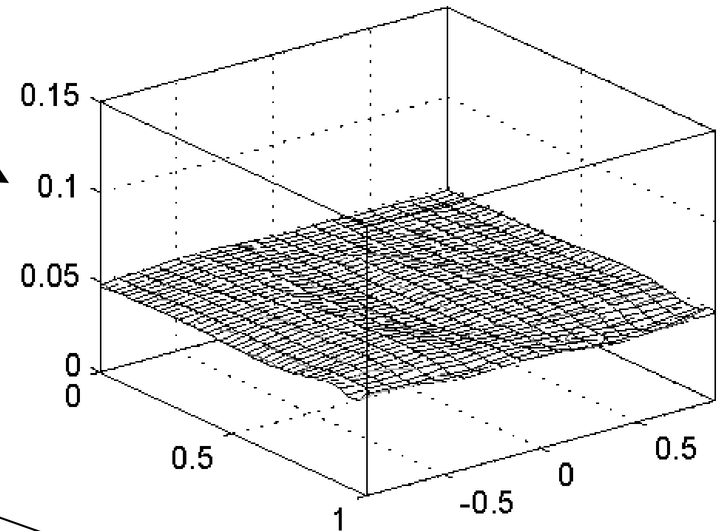
$$g(z) \equiv \tanh(x) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

Why neural nets?

- Fitting the net is a computationally demanding task. However the applied researcher has it easy. The problem is solved; our parameter values are on page 17.
- This net delivers the conditional distribution in closed form for any quantile and sample size.
- This is a type of “mixture of experts” neural net. Xiaohong Chen and Hal White showed that neural nets do a good job with high dimensional problems.

Neural net versus bootstrap

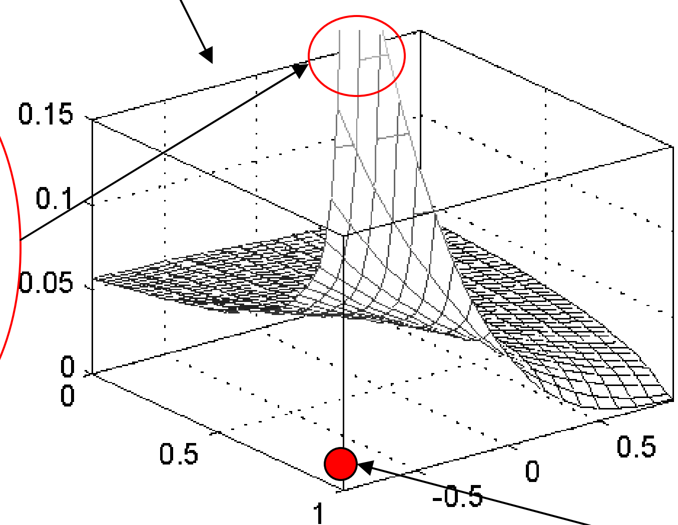
Null rejection rates using the neural network approximation to the conditional critical function.

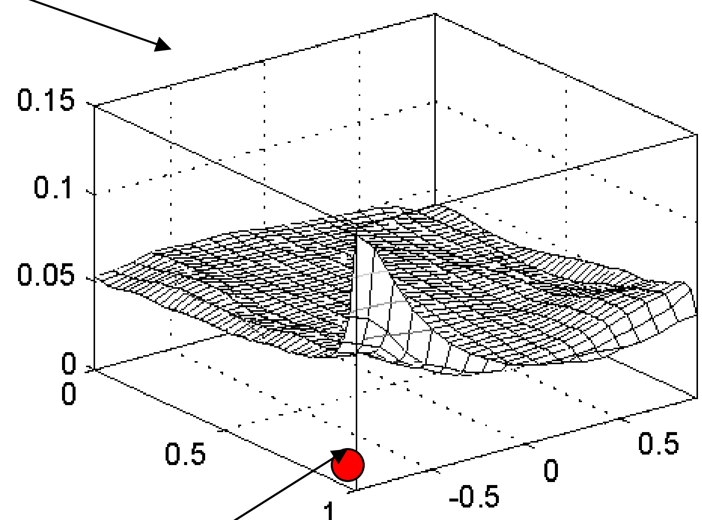


Bootstrap.

Classical asymptotics.

For $\rho=1$, it gets as high as 40%.

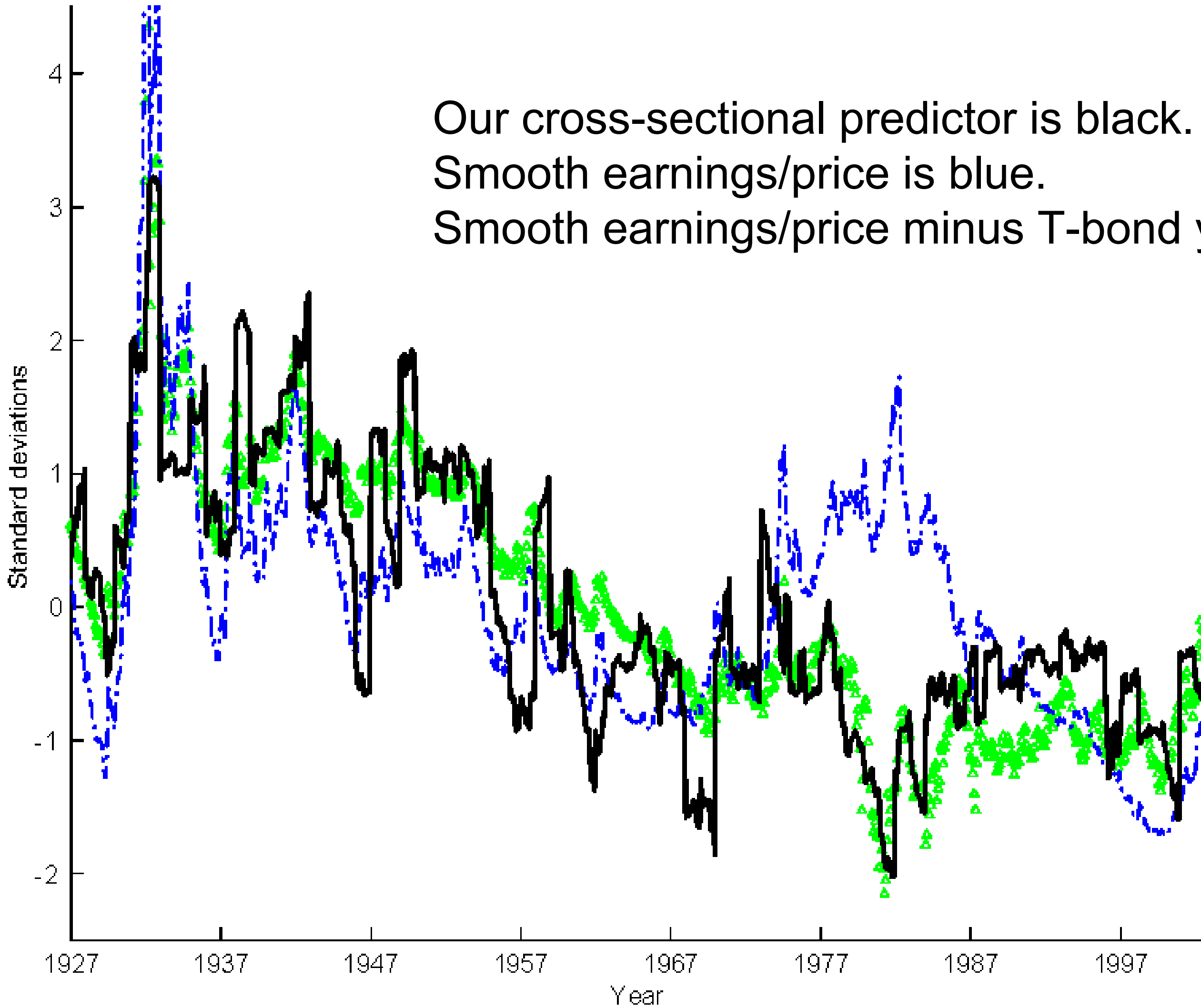




When the dividend yield is the predictor, Corr is -.9 and ρ is .97, leading to overrejection of a true null of no predictability.

EMPIRICAL RESULTS

Predictors



Our cross-sectional predictor is black.
Smooth earnings/price is blue.
Smooth earnings/price minus T-bond yield is green.

Predictive regressions

Dates	$\widehat{\theta}$	t-stat	p-value	95% conf int	$\widehat{\rho}$	$\widehat{\gamma}$	$\widehat{\sigma}_1$	$\widehat{\sigma}_2$
Prediction by the cross-sectional beta premium, $x_t = \lambda_t^{REG}$								
1927:5 - 2002:12	.0908	3.605	.000	[.042, .141]	.937	.0644	.0552	.0255
1927:5 - 1965:2	.103	3.144	.001	[.038, .168]	.934	-.0117	.0630	.0322
1965:2 - 2002:12	.0160	.297	.347	[-.086, .128]	.913	.268	.0460	.0163
1927:5 - 1946:3	.148	2.295	.014	[.018, .276]	.919	-.0475	.0820	.0333
1946:3 - 1965:2	.0817	3.186	.001	[.032, .134]	.939	.0905	.0347	.0309
1965:2 - 1984:1	.0194	.292	.348	[-.107, .159]	.911	.203	.0458	.0193
1984:1 - 2002:12	.0203	.219	.337	[-.146, .226]	.918	.385	.0462	.0126
Prediction by log smoothed earnings/price, $x_t = ep_t$								
1927:5 - 2002:12	.0170	3.454	.014	[.002, .024]	.993	-.669	.0552	.0464
1927:5 - 1965:2	.0317	3.282	.018	[.002, .046]	.987	-.671	.0630	.0549
1965:2 - 2002:12	.00756	1.319	.544	[-.011, .012]	.996	-.668	.0459	.0359
1927:5 - 1946:3	.0410	2.670	.083	[-.009, .061]	.981	-.659	.0817	.0707
1946:3 - 1965:2	.0294	2.344	.180	[-.013, .043]	.994	-.727	.0351	.0322
1965:2 - 1984:1	.0204	1.817	.341	[-.017, .029]	.987	-.662	.0455	.0362
1984:1 - 2002:12	.0105	1.251	.569	[-.019, .016]	.990	-.668	.0460	.0352

Heteroskedasticity

Dates	$\hat{\theta}$	White t-stat	p-value	95% conf int	$\hat{\rho}$	$\hat{\gamma}$	$\hat{\sigma}_1$	$\hat{\sigma}$
Prediction by the cross-sectional beta premium, $x_t = \lambda_t^{REG}$								
1927:5 - 2002:12	0.0908254	2.31335	0.0110282	[0.014, 0.169]	.93656	.0643553	.0551768	.025
1927:5 - 1965:2	0.103415	2.24512	0.0143214	[0.012, 0.195]	.934123	-.0117367	.0630215	.032
1965:2 - 2002:12	0.0160213	0.302396	0.344815	[-0.084, 0.126]	.91317	.268442	.045993	.016
1927:5 - 1946:3	0.148425	1.61165	0.058767	[-0.038, 0.33]	.918826	-.0474952	.0820265	.033
1946:3 - 1965:2	0.0817383	3.20608	0.000943039	[0.033, 0.134]	.939181	.0904511	.0347279	.030
1965:2 - 1984:1	0.0193848	0.301695	0.345021	[-0.102, 0.154]	.910814	.202914	.0458091	.019
1984:1 - 2002:12	0.0202789	0.224545	0.334831	[-0.142, 0.221]	.917989	.384513	.0462005	.012
Prediction by log smoothed earnings/price, $x_t = ep_t$								
1927:5 - 2002:12	0.0169756	2.07205	0.162478	[-0.009, 0.028]	.992709	-.669088	.0552088	.046
1927:5 - 1965:2	0.0317427	1.67639	0.252047	[-0.027, 0.059]	.986649	-.670721	.0629609	.054
1965:2 - 2002:12	0.00755646	1.17495	0.600094	[-0.014, 0.012]	.995531	-.667767	.0459093	.035
1927:5 - 1946:3	0.0409578	1.62831	0.323587	[-0.042, 0.073]	.981075	-.659176	.0816968	.070
1946:3 - 1965:2	0.0293931	2.72981	0.100753	[-0.007, 0.041]	.993735	-.726749	.0350752	.032
1965:2 - 1984:1	0.0203741	1.78598	0.35193	[-0.018, 0.03]	.987186	-.661863	.0454867	.036
1984:1 - 2002:12	0.0105404	1.18799	0.594054	[-0.02, 0.016]	.989574	-.668109	.0460461	.035

Never significant!

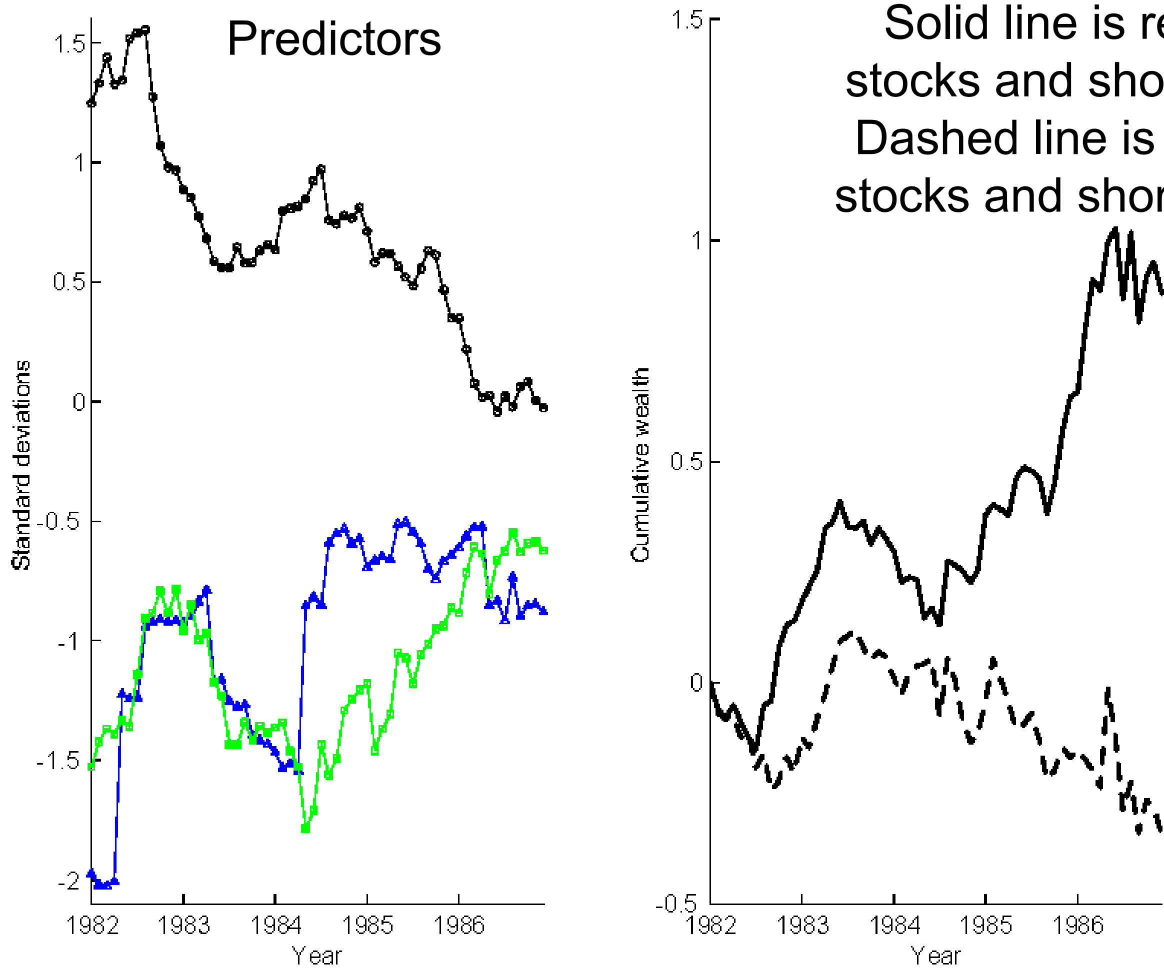
Bivariate regressions

Dates	$\widehat{\theta}_1$	t-stat	p-val	95% conf	$\widehat{\theta}_2$	t-stat	p-val	95% conf	F p-val
Prediction equation: $R^e_{M,t} = \theta_0 + \theta_1 \lambda^{SRC}_{t-1} + \theta_2 ep_{t-1} + u_t$									
1927:5 - 2002:12	.012	1.17	.158	[-.009,.031]	.013	2.32	.108	[-.004,.021]	.020
1927:5 - 1965:2	.001	.045	.532	[-.043,.041]	.031	2.17	.091	[-.010,.049]	.056
1965:2 - 2002:12	.011	.527	.285	[-.030,.057]	.008	1.36	.447	[-.009,.013]	.592
1927:5 - 1946:3	-.002	-.030	.573	[-.118,.098]	.042	1.79	.176	[-.029,.072]	.139
1946:3 - 1965:2	.037	2.03	.038	[.001,.071]	.004	.225	.798	[-.050,.023]	.086
1965:2 - 1984:1	.027	1.08	.183	[-.027,.074]	.023	2.03	.172	[-.013,.037]	.210
1984:1 - 2002:12	-.034	-.738	.778	[-.126,.057]	.012	1.39	.341	[-.014,.021]	.527
Prediction equation: $R^e_{M,t} = \theta_0 + \theta_1 \lambda^{REG}_{t-1} + \theta_2 ep_{t-1} + u_t$									
1927:5 - 2002:12	.066	2.36	.015	[.010,.124]	.012	2.12	.151	[-.005,.019]	.003
1927:5 - 1965:2	.057	1.29	.130	[-.030,.134]	.021	1.59	.241	[-.018,.037]	.031
1965:2 - 2002:12	.022	.405	.326	[-.084,.140]	.008	1.35	.445	[-.009,.013]	.615
1927:5 - 1946:3	.057	.645	.329	[-.143,.233]	.032	1.49	.272	[-.036,.062]	.123
1946:3 - 1965:2	.073	2.16	.026	[.003,.136]	.006	.397	.801	[-.043,.023]	.063
1965:2 - 1984:1	.092	1.25	.157	[-.056,.230]	.027	2.19	.101	[-.010,.044]	.160
1984:1 - 2002:12	.002	.019	.530	[-.186,.180]	.011	1.23	.415	[-.016,.020]	.623

Premia divergence in the 80's: Hypothesis 1

- Our predictor works well, but makes a gigantic wrong call in the early 1980s. Why?
- We suspect that the high equity premium in the 1980's is due to events that affected all long-term assets irrespective of their risks or betas.
- Our cross sectional predictor forecasts high returns when risky (high beta) stocks are cheap relative to safer (low beta) stocks.
- But what if all stocks are cheap? Our cross sectional measure won't work as well as a time series measure like ep.

10 year bonds versus stocks



Solid line is return from buying stocks and shorting 90 day T bills. Dashed line is return from buying stocks and shorting 10 year bonds.

Which equity premium?

- Long term bonds also did exceedingly well in the 1980s. So maybe the market's great performance was due to the general good performance of long term assets, not because of the good performance of risky assets relative to less risky ones.
- We want to predict the equity premium, the expected return on stocks over bonds. The question is, which bonds?
- In the early 1980s our measure does accurately forecast the excess return of stocks over long term bonds.
- Our measure wouldn't be useful to a portfolio manager, who moves in and out of stocks and short term Treasuries. But it would be useful to a long-term investor who buys and holds long term bonds.

Premia divergence in the 80's: Hypothesis 2

- The market made a big mistake in 1982
- The market was irrationally pessimistic
- Given this pessimism, the market priced the cross-section of stocks appropriately.

CONCLUSIONS

Conclusions

- We use the cross-sectional risk price to forecast the time series of equity premia
- Our predictor is superior to the traditional time-series predictors, except in the early 1980s
- We find similar results in the international data
- This is an example where good statistics changes the answer:
  - Classical statistics makes the valuation measures look good
  - Conditional inference gives more support to our cross-sectional predictor
- Our fitted neural networks make it very easy for the other researchers with the same problem to carry out conditional inference

Future work

- It's easy to carry out conditional inference in long-horizon regressions.
- We still do not have a completely satisfactory solution to the multivariate problem. Possible directions include:
  - Re-parameterizing the model so that the critical function is smoother in the parameters.
  - Studying the distribution theory to reduce the effective dimensionality.
  - Analytic approximations.

Our fitting algorithm is slightly original

- We pick (Ψ,ξ) to minimize a sum of squared size distortions:

  $$(\hat{\psi},\hat{\xi}) = \arg\min \sum_{j=1}^{J}\sum_{i=1}^{N}\left(\alpha_j - B^{-1}\sum_{b=1}^{B} 1_h\left(\hat{t}_{b,i} > q_{\alpha_j}^{nn}\left(X_{b,i},\psi,\xi\right)\right)\right)^2$$

  Compare this with the objective function in White 1992, Koenker and Hendricks 1992 and Doksum and Koo 2000:

  $$\sum_{i=1}^{N} Q_\alpha\left(\hat{t}_{b,i} - q_\alpha^{nn}(X_{bi},\psi,\xi)\right); \quad \mathrm{Q}_\alpha(x) = \alpha|x|1(x \geq 0) + (1-\alpha)|x|1(x < 0)$$

- Q_α is the usual check-function used in quantile regression. Koenker and Portnoy (1997) have a nice algorithm for minimizing Q_α when the conditional quantile is linear.
- Why is our function better? It’s the “loss function” used by applied researchers running Monte Carlos. They don’t care about the quantile per se. They want small size distortions.

Measure λ^{SRC}

- $VALRANK_{i,t}$ is the valuation rank of firm i at time t:
  - Each year we rank stocks on D/P, BE/ME, E/P, and CF/P
  - VALRANK is the average of (up to) four percentile ranks
- We estimate CAPM betas for all stocks with up to three years of past monthly data
- $\lambda_t^{SRC} \equiv$ cross-sectional Spearman rank correlation of past $\beta_{i,t}$ and $VALRANK_{i,t}$

Measure λ^{REG}

- $GROWTHRANK^{g}_{i,t}$ are our growth proxies:
  - Each year we rank stocks on D/BE, non-dividend-paying dummy, long-term ROE, transitory ROE, loss dummy, industry concentration
  - $GROWTHRANK^{g}_{i,t}$ is the normalized rank on g^{th} variable, highest value gets rank 1 and lowest value gets rank 0

$$\lambda_t^{REG} \equiv \text{Cross-sectional regression coefficient in the regression}$$

$$VALRANK_{i,t} = \lambda_{0,t} + \lambda_t^{REG}\,\hat{\beta}_{i,t} + \sum_{g=1}^{6} \lambda_t^{g}\, GROWTHRANK_{i,t} + \varepsilon_{i,t}$$

Measure λ^{MSCI}

- Take the top-30% and bottom-30% portfolios sorted on four of Morgan Stanley Capital International's value measures: D/P, BE/ME, E/P, and C/P.
- Estimate the local-market betas for these portfolios using a three-year rolling window
- λ^{MSCI} is the average beta of the four value portfolios minus the average beta of the four growth portfolios.
- Available for an international sample of 22 countries

Measures λ^{DP} and λ^{DPG}

- Sort stocks into five portfolios on the end-of-May dividend yield
- For each portfolio measure value-weight average dividend yield and the value-weight average of stock-level rolling betas
- Regress these five portfolio-level dividend yields in levels on the portfolios' betas in a simple regression to produce λ^{DP}
- λ^{DPG} is the multiple regression coefficient of the portfolio's dividend yield on its beta, controlling for its value-weight one-year dividend growth

Measures λ^{BM} and λ^{BMG}

- Sort stocks into five portfolios on the end-of-May BE/ME
- For each portfolio measure value-weight average BE/ME and the value-weight average of stock-level rolling betas
- Regress these five portfolio-level BE/ME in levels on the portfolios' betas in a simple regression to produce λ^{BM}
- λ^{BMG} is the multiple regression coefficient of the portfolio's BE/ME on its beta, controlling for its value-weight one-year ROE

Measure λ^{ER}

- Each month, we regress cross-sectionally demeaned firm-level returns on lagged cross-sectionally demeaned characteristics using a ten-year window:
  - VALRANK (as computed for construction of λ^{SRC})
  - Rolling beta estimates β
  - Valuation multiples D/P, BE/ME, E/P, and C/P in levels
  - Profitability and growth controls used for λ^{REG} in levels
- In a cross-sectional regression, regress the current forecasts on rolling betas and define λ^{ER} as the regression coefficient