Present Value and Predictable Returns (3)

John Y. Campbell

Ec2723

September 2010

Outline

- Loglinear present value models.
  - Campbell-Shiller return approximation
  - VAR approach
  - Illustrative model with AR(1) expected return
- Predictive return regressions
  - Long vs. short horizons
- Persistent regressor problem
  - Stambaugh bias
  - Recent responses

Loglinear Present Value Models

- Last time we discussed linear present value models.
- These have difficulty capturing time-varying discount rates, except in special cases
  - Linearity generating process with "twisted" AR(1)
  - Drifting steady state model with random walk for log D/P
- Loglinear approximation is an alternative way to capture the price effects of time-varying discount rates.

Campbell-Shiller Return Approximation

$$
\begin{aligned}
r_{t+1} &= \log(1+R_{t+1}) = \log(P_{t+1}+D_{t+1}) - \log(P_t) \\
&= p_{t+1} - p_t + \log(1+\exp(d_{t+1}-p_{t+1})).
\end{aligned}
$$

Approximate the nonlinear function

$$\log(1+\exp(d_{t+1}-p_{t+1})) = f(d_{t+1}-p_{t+1})$$

as

$$f(d_{t+1}-p_{t+1}) \approx f(\overline{d-p}) + f'(\overline{d-p})(d_{t+1}-p_{t+1}-(\overline{d-p})).$$

Here $f(x) = \log(1+\exp(x))$ and $f'(x) = \exp(x)/(1+\exp(x))$.

Campbell-Shiller Return Approximation

We get

$$r_{t+1} \approx k + \rho p_{t+1} + (1-\rho) d_{t+1} - p_t,$$

where

$$\rho = \frac{1}{1 + \exp(\overline{d-p})},$$

and

$$k = -\log(\rho) - (1-\rho)\log(1/\rho - 1).$$

- Replace the log of a sum with an average of logs, where the relative weights depend on the average relative magnitudes of dividend and price.

Price Implications

$$r_{t+1} \approx k + \rho p_{t+1} + (1-\rho) d_{t+1} - p_t,$$

Solve difference equation forward, imposing terminal condition

$$\lim_{j \longrightarrow \infty} \rho^j p_{t+j} = 0.$$

We get

$$p_t = \frac{k}{1-\rho} + \sum_{j=0}^{\infty} \rho^j [(1-\rho) d_{t+1+j} - r_{t+1+j}].$$

- This is an approximate accounting identity. It holds ex post.
- So it should hold in expectation, not just for RE but for all expectations that respect identities.

Price Implications

$$p_t = \frac{k}{1-\rho} + \mathrm{E}_t \sum_{j=0}^{\infty} \rho^j \left[(1-\rho) d_{t+1+j} - r_{t+1+j} \right] = \frac{k}{1-\rho} + p_{CF,t} - p_{DR,t}.$$

- $p_{CF,t}$ is the component of the price due to cash flow (dividend) expectations
- $p_{DR,t}$ is the component due to discount rate (return) expectations.

What if log dividends follow a unit root process? Then we can subtract d_t from both sides:

$$d_t - p_t = \frac{-k}{1-\rho} + \mathrm{E}_t \sum_{j=0}^{\infty} \rho^j \left[-\Delta d_{t+1+j} + r_{t+1+j} \right].$$

- $d_t - p_t$ is stationary, so log dividends and prices are cointegrated, with a known cointegrating vector.

An Earnings-Based Approach

Vuolteenaho (2002):

$$b_t - v_t = \mu + E_t \sum_{j=0}^{\infty} \rho^j \left[-roe_{t+1+j} + r_{t+1+j} \right],$$

where v_t is the log market value of the firm and $roe_t = \log(1 + ROE_t)$.

- This works well for studying individual firms that may not have a stable dividend policy.

Return Implications

Substitute price solution back into return approximation:

$$
\begin{aligned}
r_{t+1} - \mathrm{E}_t r_{t+1} &= (\mathrm{E}_{t+1} - \mathrm{E}_t) \sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} - (\mathrm{E}_{t+1} - \mathrm{E}_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j} \\
&= N_{CF,t+1} - N_{DR,t+1}.
\end{aligned}
$$

- $N_{CF,t}$ is the revision in expectations (news) about current and future cash flows. (Sum starts at 0.)
- $N_{DR,t}$ is the revision in expectations (news) about future discount rates. (Sum starts at 1.)
- Surprising implication: Better information about future dividends lowers the volatility of returns.

VAR Approach

$$x_{t+1} = Ax_t + \epsilon_{t+1}$$

for a vector x_t with first element equal to return.

- First-order VAR assumption not restrictive because higher-order VAR can be rewritten in this form with an expanded state vector and a singular variance-covariance matrix of innovations.

Then

$$\mathrm{E}_t x_{t+1+j} = A^{j+1} x_t.$$

$$r_{t+1} - \mathrm{E}_t r_{t+1} = e1' \epsilon_{t+1},$$

where $e1' = [10...0]$.

VAR Approach

$$N_{DR,t+1} = e1' \sum_{j=1}^{\infty} \rho^j A^j \epsilon_{t+1} = e1' \rho A (I - \rho A)^{-1} \epsilon_{t+1},$$

and

$$N_{CF,t+1} = r_{t+1} - E_t r_{t+1} + N_{DR,t+1} = e1'(I + \rho A (I - \rho A)^{-1}) \epsilon_{t+1}.$$

- Decomposition is invariant to the inclusion of return rather than dividend growth, provided that the system includes $d - p$. Why?
- Decomposition is empirically insensitive to the inclusion of return rather than dividend growth, if some other persistent valuation ratio is included.
- Decomposition is sensitive to the information variables in the VAR.

An Illustrative Model

- Warning: Change in notation! Scalar variable x_t drives expected return.
- x_t is an AR(1).

$$\mathrm{E}_t r_{t+1} = r + x_t,$$

$$x_{t+1} = \phi x_t + \xi_{t+1.}$$

$$r_{t+1} = r + x_t + u_{t+1}.$$

Price Implications

$$p_{DR,t} = \mathrm{E}_t \sum_{j=0}^{\infty} \rho^j r_{t+1+j} = \frac{r}{1-\rho} + \frac{x_t}{1-\rho\phi}.$$

$$\mathrm{Var}(p_{DR,t}) = \frac{\sigma_x^2}{(1-\rho\phi)^2},$$

- The expected return may have a very small volatility yet may still have a very large effect on the stock price if it is highly persistent.

$$N_{DR,t+1} = (\mathrm{E}_{t+1} - \mathrm{E}_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j} = \frac{\rho \xi_{t+1}}{1-\rho\phi} \approx \frac{\xi_{t+1}}{1-\phi}.$$

- A 1% increase in the expected return today is associated with a capital loss of about 2% if the AR coefficient is 0.5, a loss of about 4% if the AR coefficient is 0.75, and a loss of about 10% if the AR coefficient is 0.9.

Back to Bob Shiller

This illustrates Bob Shiller's point: "Returns on speculative assets are nearly unforecastable; this fact is the basis of the most important argument in the oral tradition against a role for mass psychology in speculative markets. One form of this argument claims that because real returns are nearly unforecastable, the real price of stocks is close to the intrinsic value, that is, the present value with constant discount rate of optimally forecasted future real dividends. This argument... is one of the most remarkable errors in the history of economic thought".

Discussion: What has the global financial crisis done to the reputation of the efficient market hypothesis?

Return Autocovariance Implications

$$\gamma_i = \mathrm{Cov}(r_{t+1}, r_{t+1+i}) = \phi^{i-1}\left[C + \left(\frac{\phi}{1-\phi^2} - \frac{\rho}{1-\rho\phi}\right)\sigma_\xi^2\right],$$

where

$$C = \mathrm{Cov}(\xi_{t+1}, N_{CF,t+1}).$$

- Autocovariances are all of the same sign and die off at rate ϕ.
- Sign of autocovariances depends on three terms. How to interpret them?
- Autocovariances can all be zero, even if expected returns vary through time. This shows that prices can be weak-form efficient even if they are not semi-strong form efficient.
- However for reasonable parameter values (C not strongly positive, ϕ not too large), autocorrelations will tend to be negative.

Predictive Return Regressions

For simplicity, assume $C = 0$. Then the variance of the stock return is

$$\mathrm{Var}(r_{t+1}) = \sigma^2_{CF} + \sigma^2_x \frac{1+\rho^2 - 2\rho\phi}{(1-\rho\phi)^2} \approx \sigma^2_{CF} + \frac{2\sigma^2_x}{1-\phi},$$

where $\sigma^2_{CF} = \mathrm{Var}(N_{CF})$.
The R^2 of a single-period return regression onto x_t is

$$\begin{aligned} R^2(1) &= \frac{\mathrm{Var}(E_t r_{t+1})}{\mathrm{Var}(r_{t+1})} \approx \frac{\sigma^2_x}{\sigma^2_{CF} + 2\sigma^2_x/(1-\phi)} \\ &= \left(\frac{\sigma^2_{CF}}{\sigma^2_x} + \frac{2}{1-\phi} \right)^{-1} \leq \frac{1-\phi}{2}. \end{aligned}$$

When x_t is extremely persistent, the one-period return regression must have a low R^2, even if there is no cash flow news at all!

Long-Horizon Return Regressions

$$r_{t+1} + ... + r_{t+K} = \beta(K) x_t,$$

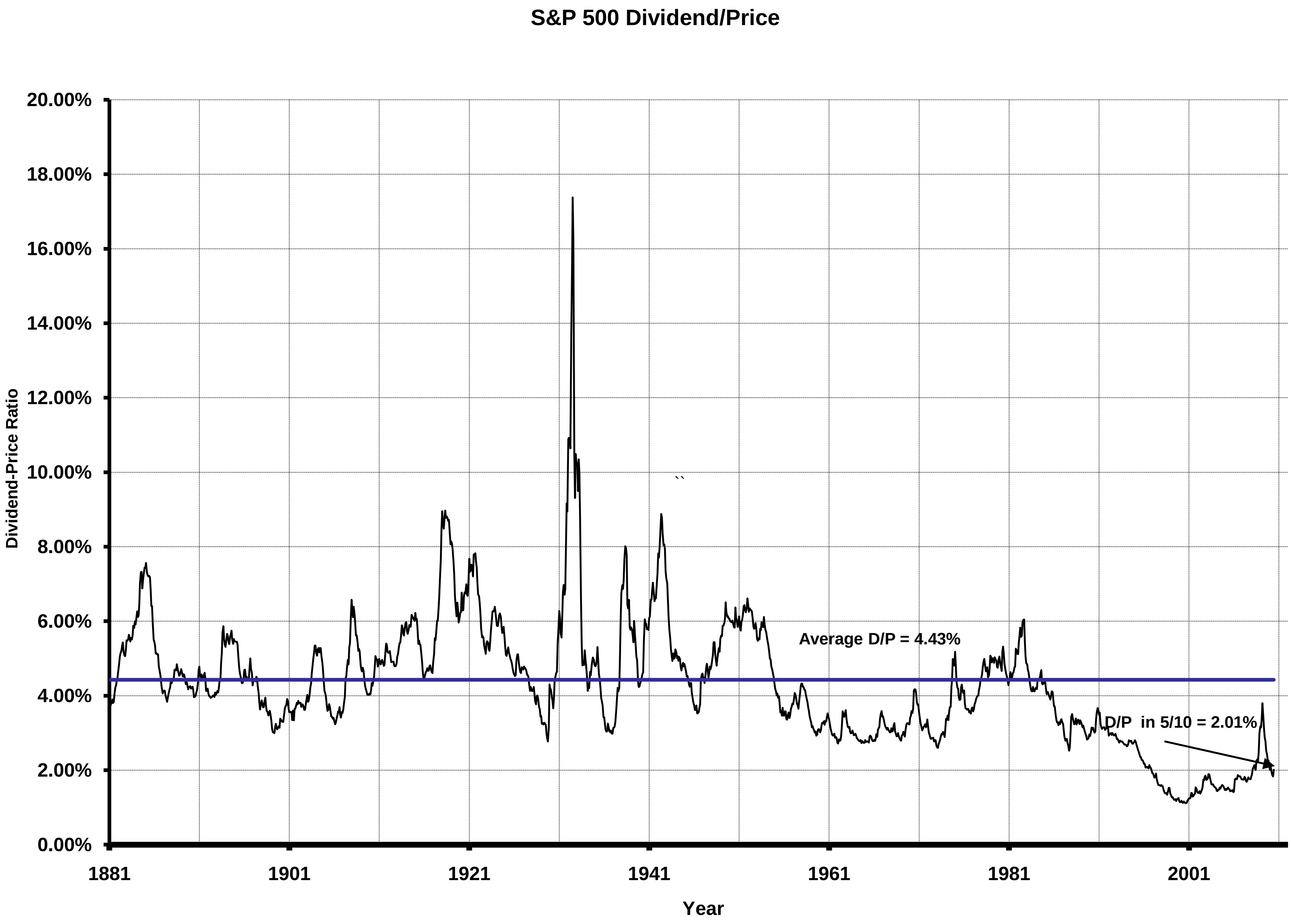
where

$$\beta(K) = 1 + \phi + ... + \phi^{K-1} = \frac{1-\phi^K}{1-\phi}.$$

The ratio of the K-period R^2 to the 1-period R^2 is

$$\begin{aligned}
\frac{R^2(K)}{R^2(1)} &= \left[\frac{\mathrm{Var}(E_t r_{t+1} + ... + E_t r_{t+K})}{\mathrm{Var}(r_{t+1} + ... + r_{t+K})}\right] / \left[\frac{\mathrm{Var}(E_t r_{t+1})}{\mathrm{Var}(r_{t+1})}\right] \\
&= \frac{\beta(K)^2}{\beta(1)^2} \frac{1}{KV(K)} = \left(\frac{1-\phi^K}{1-\phi}\right)^2 \frac{1}{KV(K)}.
\end{aligned}$$

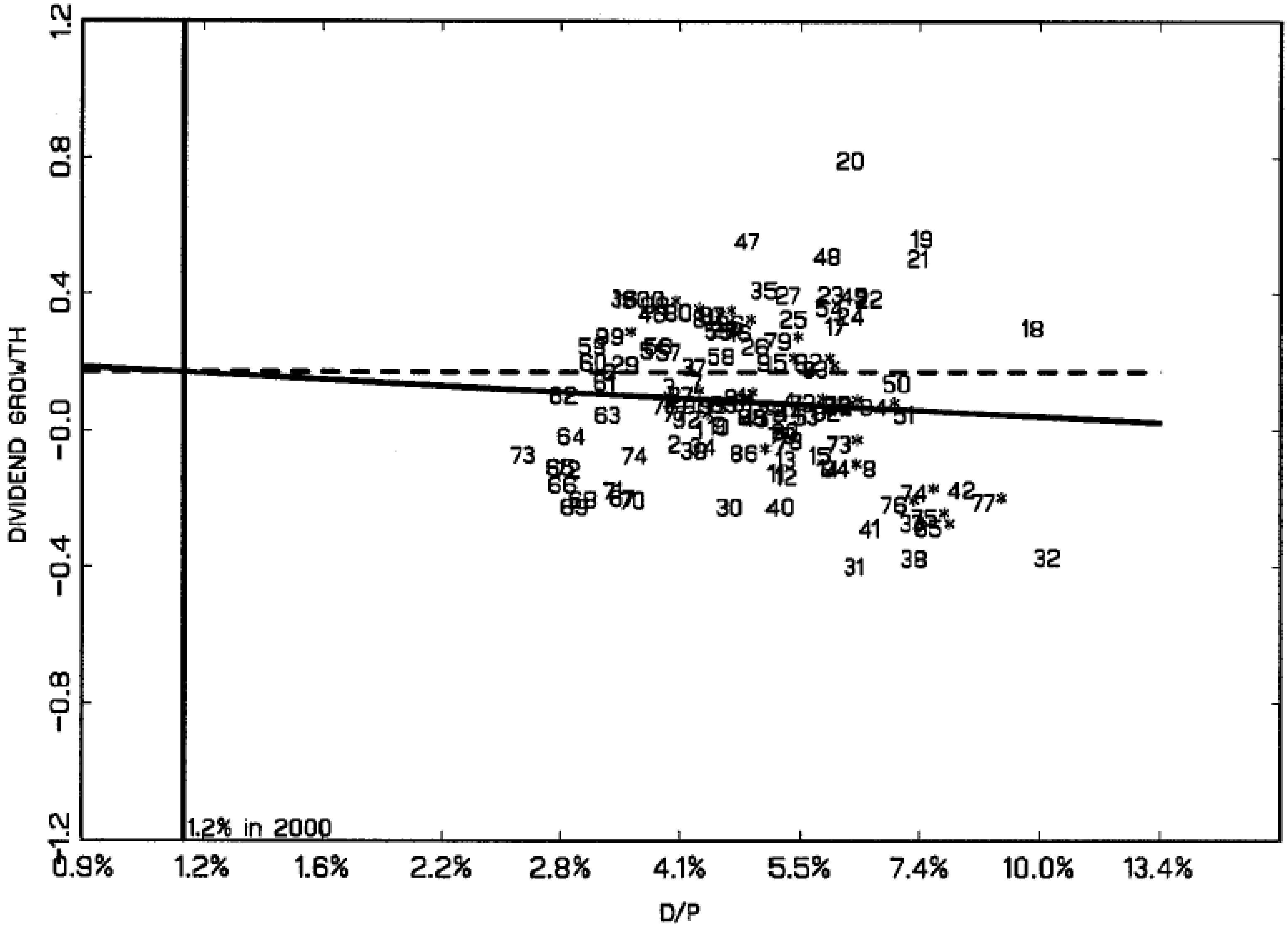
This grows at first with K if ϕ is large, then eventually dies away to zero.



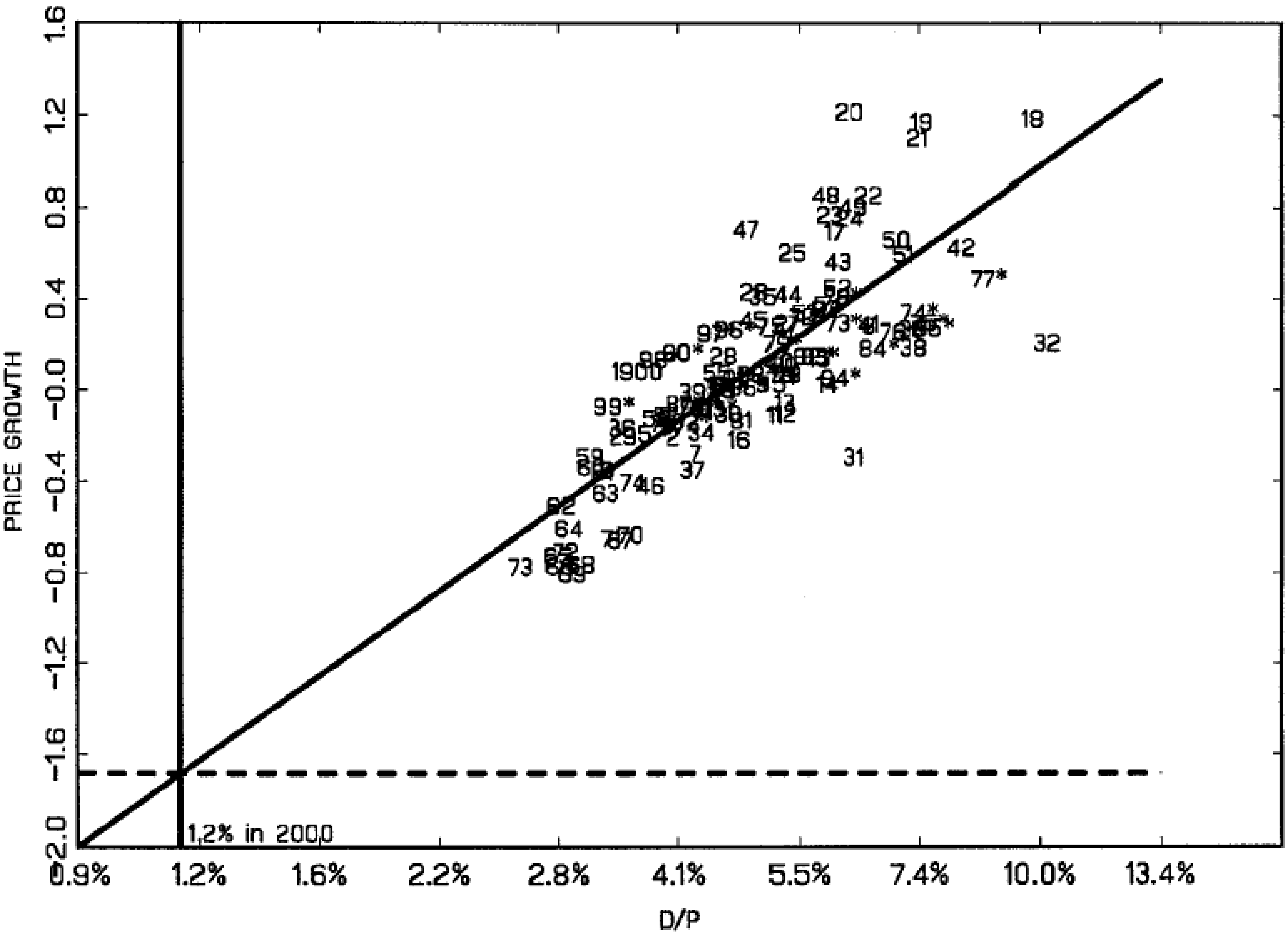
S&P 500 Dividend/Price
Dividend-Price Ratio
20.00%
18.00%
16.00%
14.00%
12.00%
10.00%
8.00%
6.00%
4.00%
2.00%
0.00%
Average D/P = 4.43%
D/P in 5/10 = 2.01%
1881
1901
1921
1941
1961
1981
2001
Year

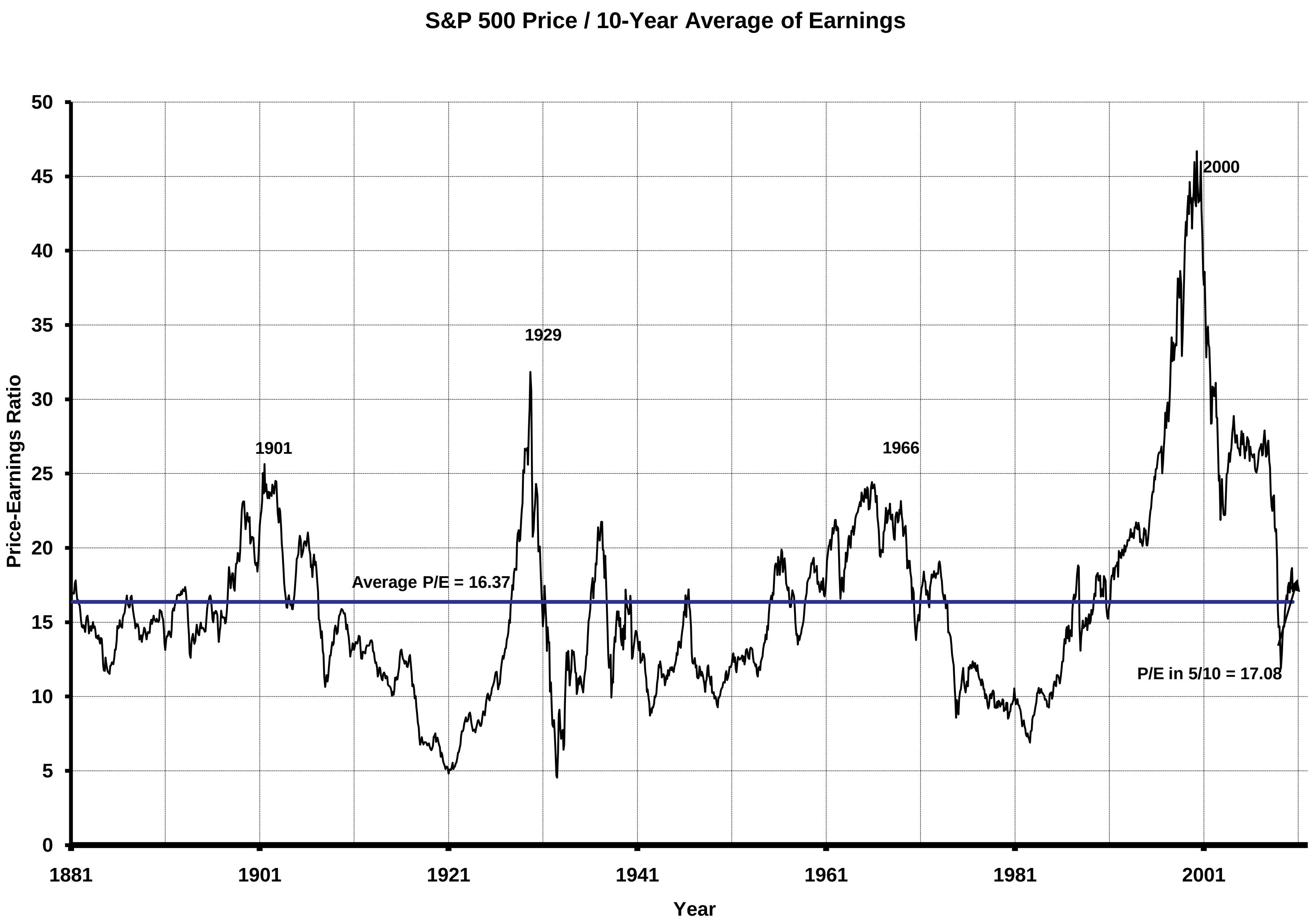
Campbell and Shiller 2005

Figure 1. DIVIDEND GROWTH till the next time D/P crosses its mean

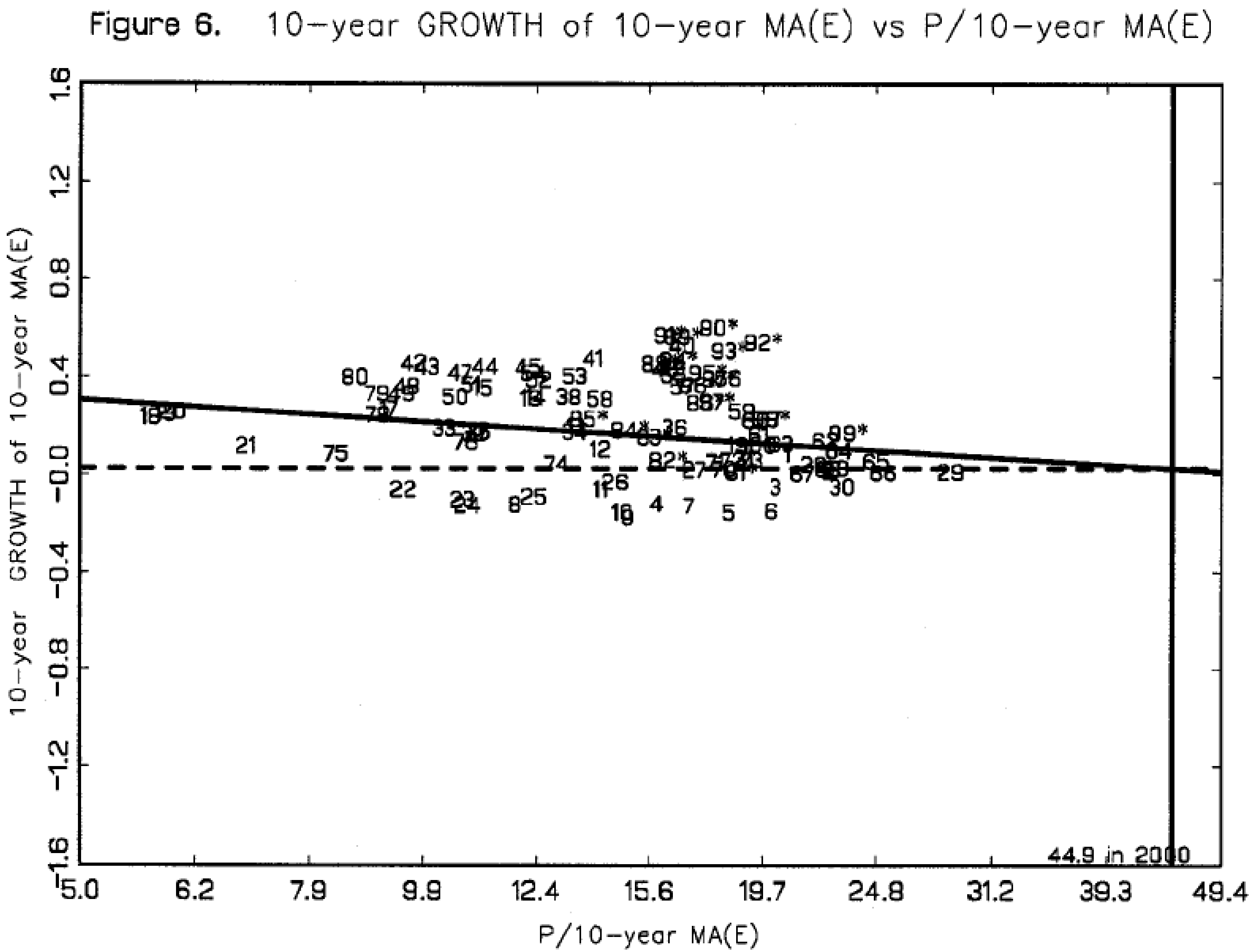


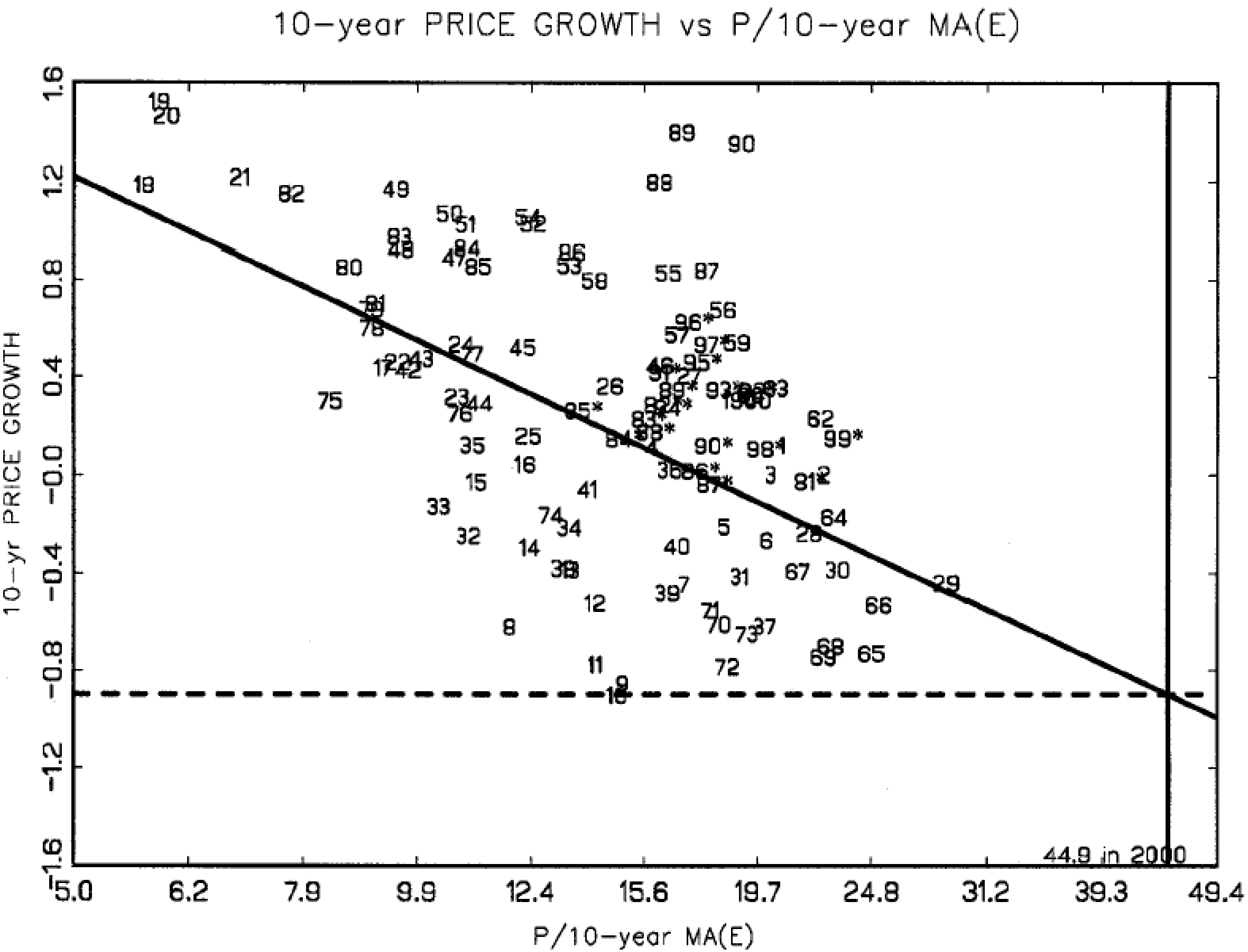
PRICE GROWTH till the next time D/P crosses its mean




S&P 500 Price / 10-Year Average of Earnings
Price-Earnings Ratio
0
5
10
15
20
25
30
35
40
45
50
1881
1901
1921
1941
1961
1981
2001
Year
1901
1929
1966
2000
Average P/E = 16.37
P/E in 5/10 = 17.08

Campbell and Shiller 2005





10-year PRICE GROWTH vs P/10-year MA(E)
10-yr PRICE GROWTH
P/10-year MA(E)
44.9 in 2000

Long-Horizon Return Regressions

- Much higher R^2 statistics than short-horizon regressions.
- The usual asymptotic t-statistics (with Newey-West correction for overlapping observations) deliver stronger rejections.
- However these t-statistics tend to have size distortions when the overlap is large relative to the sample size, so the long-horizon regression evidence is tenuous statistically.
- Pastor and Stambaugh (2009) have recently argued for the use of a "predictive system", in which the AR(1) model for expected return is combined with a vector of return predictors that are used to deliver filtered estimates of the unobservable expected return.

The Persistent Regressor Problem

Kendall (1954): finite-sample bias in estimation of AR(1),

$$x_{t+1} = \phi x_t + \xi_{t+1},$$

is

$$\mathrm{E}[\widehat{\phi} - \phi] = -\left(\frac{1+3\phi}{T}\right) + o\left(\frac{1}{T^2}\right).$$

- This bias arises primarily because the mean of x is unknown and must be estimated.

Stambaugh Bias

Stambaugh (1999): finite-sample bias in one-period predictive regression

$$\mathrm{E}[\widehat{\beta}-\beta]=\gamma\mathrm{E}[\widehat{\phi}-\phi],$$

where $\gamma=\sigma_{u\xi}/\sigma_{\xi}^{2}$, the regression coefficient of the innovation to return on the innovation to the predictor variable.

In the case where the dividend-price ratio is the predictor variable, we expect $\gamma<0$ so downward bias in $\widehat{\phi}$ produces upward bias in $\widehat{\beta}$:

$$\mathrm{E}[\widehat{\beta}-\beta]=-\gamma\left(\frac{1+3\phi}{T}\right)=\frac{\rho(1+3\phi)}{(1-\rho\phi)T},$$

where the second equality holds for the case where $C=0$.

Table 1
Finite-sample properties of $\hat{\beta}$

The table reports finite-sample properties of the ordinary least squares (OLS) estimator $\hat{\beta}$ in the regression

$$y_t = \alpha + \beta x_{t-1} + u_t.$$

The sampling properties are computed under the assumption that x_t obeys the process

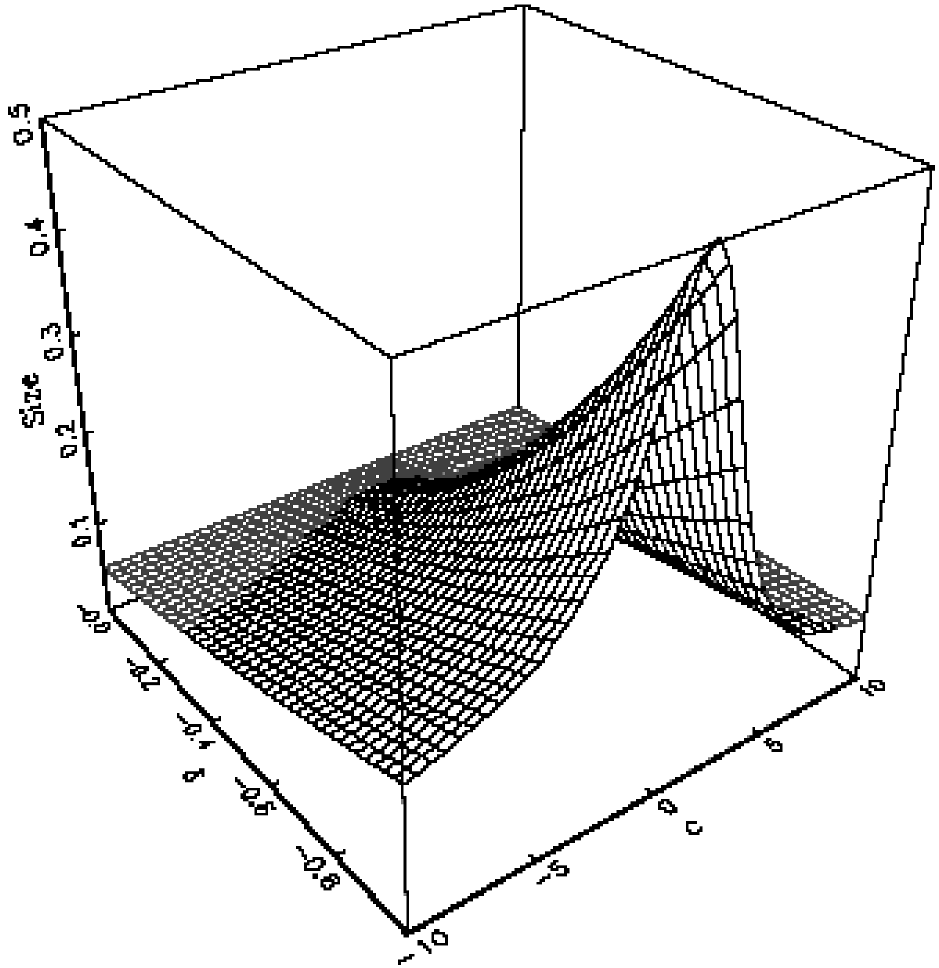
$$x_t = \theta + \rho x_{t-1} + v_t,$$

where $\rho^2 < 1$ and $[u_t\ v_t]'$ is distributed $\mathrm{N}(0, \Sigma)$, identically and independently across t. The true bias and higher-order moments depend on ρ and Σ (with distinct elements σ_u^2, σ_v^2, and σ_{uv}). For each sample period, those parameters are set equal to the estimates obtained when y_t is the continuously compounded return in month t on the value-weighted NYSE portfolio, in excess of the one-month T-bill return, and x_t is the dividend–price ratio on the value-weighted NYSE portfolio at the end of month t. The moments in the standard setting are conditioned on $x_0, \ldots, x_{T-1}$ and ignore any dependence of u_t on those values. The p-values are associated with a test of $\beta = 0$ versus $\beta > 0$

	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
C. Sample characteristics and parameter values				
$\hat{\beta}$	0.21	0.21	0.44	0.19
T	840	300	540	240
ρ	0.972	0.948	0.980	0.987
$\sigma_u^2 \times 10^4$	30.05	54.46	16.42	17.50
$\sigma_v^2 \times 10^4$	0.108	0.247	0.029	0.033
$\sigma_{uv} \times 10^4$	-1.621	-3.360	-0.651	-0.715

Stambaugh Bias

- There are similar problems with the distribution of the t-statistic when ϕ is close to one (Cavanagh and Stock).
- No problem when persistent regressor has innovations orthogonal to asset returns (e.g. inflation, interest rates predicting stock returns).
- Reverse problem for case of excess bond returns regressed on yield spread. Coefficient is biased downwards rather than upwards. Why?



Campbell and Yogo,
JFE 2006

Figure 2: Asymptotic Size of the One-Sided t-test at 5% Significance

This figure plots the actual size of the nominal 5% t-test when the largest autoregressive root of the predictor variable is $\rho = 1 + c/T$. The null hypothesis is $\beta = \beta_0$ against the one-sided alternative $\beta > \beta_0$. δ is the correlation between the innovations to returns and the predictor variable. The dark shade indicates regions where the size is greater than 7.5%.

Lewellen Response

Lewellen (2004): Condition on estimated persistence $\widehat{\phi}$ and true persistence ϕ:

$$\mathrm{E}[\widehat{\beta} - \beta \,|\, \widehat{\phi}, \phi] = \gamma[\widehat{\phi} - \phi].$$

- We do not know true ϕ, but Lewellen argues we know $\phi \leq 1$ and worst case is $\phi = 1$.
- He proposes the conservative approach of adjusting the estimated coefficient using this worst-case bias:

$$\widehat{\beta}_{adj} = \widehat{\beta} - \gamma(\widehat{\phi} - 1).$$

- Adjusted coefficient has variance σ_v^2/σ_x^2, where v is the residual in the regression of u on ξ: $u = \gamma\xi + v$.

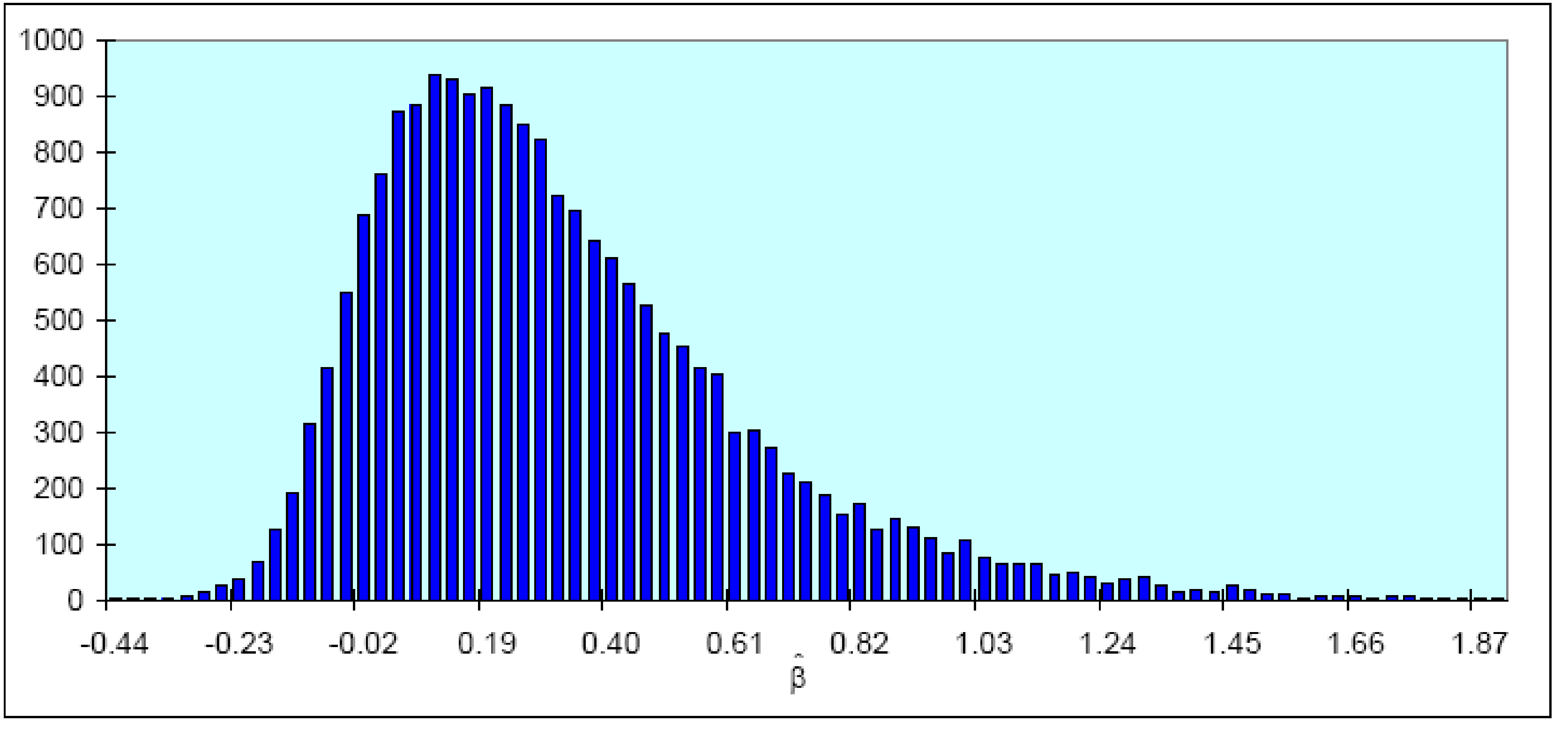
Figure 1

Lewellen, JFE 2004

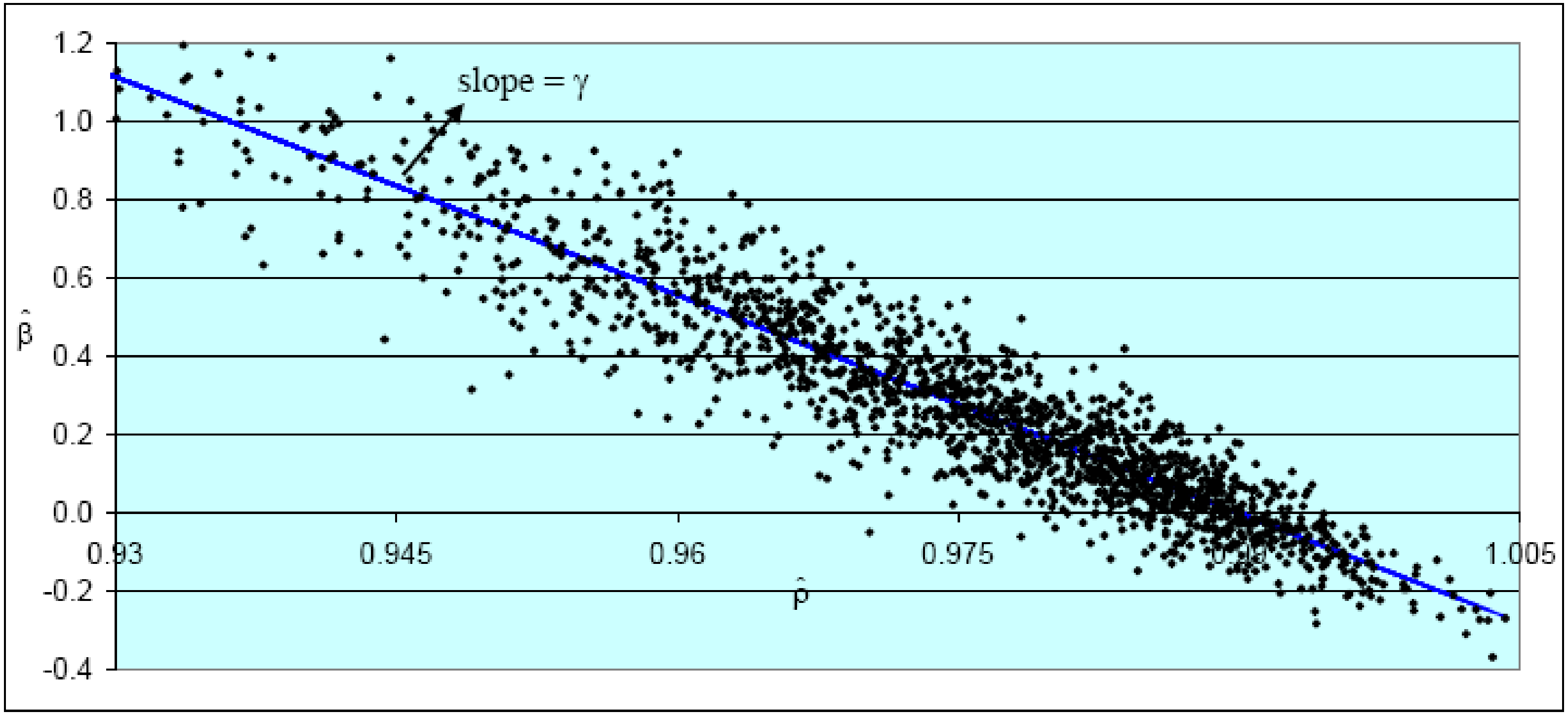
Sampling distribution of $\hat{\beta}$ and $\hat{\rho}$

The figure shows the distribution of the OLS slope estimates from $r_t = \alpha + \beta x_{t-1} + \varepsilon_t$ and $x_t = \phi + \rho x_{t-1} + \mu_t$. Panel A shows the marginal, or unconditional, distribution of $\hat{\beta}$ and Panel B shows the joint distribution of $\hat{\beta}$ and $\hat{\rho}$. The plots are based on Monte Carlo simulations (20,000 in Panel A and 2,000 in Panel B). The true parameters are $\beta = 0$, $\rho = 0.99$, $\text{cor}(\varepsilon, \mu) = -0.92$, $\sigma_\varepsilon = 0.04$, $\sigma_\mu = 0.002$, and $T = 300$.

Panel A: Marginal distribution of $\hat{\beta}$



Panel B: Joint distribution of $\hat{\beta}$ and $\hat{\rho}$



Cochrane Response

Cochrane (2008): Use the fact that $d_t - p_t$ does not forecast dividend growth, "the dog that did not bark".

- Campbell-Shiller approximation implies that if we regress r_{t+1}, Δd_{t+1}, and $d_{t+1} - p_{t+1}$ onto $d_t - p_t$, the coefficients β, β_d, and ϕ are related by
$$\beta = 1 - \rho\phi + \beta_d.$$
- If we have prior knowledge about ϕ, then β and β_d are linked. For example, if $\rho = 0.96$ and we know that $\phi \leq 1$, then $\beta_d \leq \beta - 0.04$. If $\beta = 0$, then β_d must be negative and less than -0.04.
- If the dividend-price ratio fails to predict stock returns, it will be explosive unless it predicts dividend growth. Since the dividend-price ratio cannot be explosive, the absence of predictable dividend growth strengthens the evidence for predictable returns.

Campbell-Yogo Response

Campbell-Yogo (2006): If we knew persistence, we could reduce noise by adding the innovation to the predictor variable to the predictive regression, estimating

$$r_{t+1} = \alpha' + \beta x_t + \gamma(x_{t+1} - \phi x_t) + v_{t+1}.$$

- The additional regressor, $(x_{t+1} - \phi x_t) = \eta_{t+1}$, is uncorrelated with the original regressor x_t but correlated with the dependent variable r_{t+1}. Thus we still get a consistent estimate of the original predictive coefficient β, but with increased precision because we have controlled for some of the noise in unexpected stock returns.
- Of course, in practice we do not know the persistence coefficient ϕ, but we can construct a confidence interval for it by inverting a unit root test.
- The test delivers particularly strong evidence for predictability if we rule out a persistence coefficient $\phi > 1$ on prior grounds.