Predictive Regressions with Persistent Regressors

Jan Philipp Wöltjen

June 25, 2019

Outline

- The regression problem
- Local to unity asymptotics
- Constructing a pretest
- Motivating more efficient tests
- Constructing the test statistic
- Making the test feasible
- Analysis of the power gain
- Empirical results

The Regression Setup

- Campbell and Yogo (2006)[1] (CY) consider the system of equations:
-

$$r_t = \alpha + \beta x_{t-1} + u_t \tag{1}$$
$$x_t = \gamma + \rho x_{t-1} + e_t \tag{2}$$

- For the sake of simplicity they further assume normality:

$$w_t = (u_t, e_t)' \sim \mathrm{N}(0, \Sigma), \text{ where } \Sigma \text{ is known}$$
$$\Sigma = \begin{bmatrix} \sigma_u^2 & \sigma_{ue} \\ \sigma_{\text{ue}} & \sigma_e^2 \end{bmatrix}$$

[1] Campbell, J. Y. and M. Yogo (2006). Efficient tests of stock return predictability. Journal of Financial Economics 81, 27-60.

Testing for Significance

- To test β for significance, CY start by considering the maximum likelihood ratio test.

$$\max_{\beta,\rho,\alpha,\gamma} L(\beta,\rho,\alpha,\gamma) - \max_{\rho,\alpha,\gamma} L\left(\beta_0,\rho,\alpha,\gamma\right) = t\left(\beta_0\right)^2 > C,$$

- where C is some constant and the joint log likelihood (ignoring two constants) is given by

$$\begin{aligned} L(\beta,\rho,\alpha,\gamma) = & -\frac{1}{1-\delta^2}\sum_{t=1}^{T}\left[\frac{\left(r_t-\alpha-\beta x_{t-1}\right)^2}{\sigma_u^2}\right. \\ & -2\delta\frac{\left(r_t-\alpha-\beta x_{t-1}\right)\left(x_t-\gamma-\rho x_{t-1}\right)}{\sigma_u\sigma_e} \\ & \left.+\frac{\left(x_t-\gamma-\rho x_{t-1}\right)^2}{\sigma_e^2}\right] \end{aligned} \tag{3}$$

The T-Test Ignores Information

- The LRT from above turns out to be the same when considering only the marginal likelihood

$$L(\beta, \alpha) = -\sum_{t=1}^{T} (r_t - \alpha - \beta x_{t-1})^2 \tag{4}$$

.

- It thus ignores information contained in the system.

Local to Unity Asymptotics

- The largest autoregressive root is modeled as $\rho = 1 + c/T$ where c is a fixed constant.
- This does not imply that ρ literally follows this process in practice. It is merely a tool to model autoregressive roots very close to one. It has some nice properties:
- Asymptotic distribution theory is not discontinuous when x_t is $I(1)$.

Pretesting for Size Distortions of the T-Test

- Under local-to-unity asymptotic theory the t-statistic does not converge to a standard normal distribution under the null but to functionals of a diffusion process.
- Under the null it convergences to:

$$t\left(\beta_0\right) \Rightarrow \delta \frac{\tau_c}{\kappa_c} + \left(1 - \delta^2\right)^{1/2} Z, \tag{5}$$

- where

$$\kappa_c = \left(\int J_c^{\mu}(s)^2 ds \right)^{1/2}, \tau_c = \int J_c^{\mu}(s) dW_e(s),$$

and $Z \sim \mathrm{N}(0,1)$ independent of $\left(W_e(s), J_c(s)\right)$

- $(W_u(s), W_e(s))'$ is a two-dimensional Wiener process with correlation δ.
- $J_c(s)$ solving $\mathrm{d}J_c(s) = cJ_c(s)\mathrm{d}s + \mathrm{d}W_e(s)$ with initial condition $J_c(0) = 0$.
- $J_c^{\mu}(s) = J_c(s) - \int J_c(r)\mathrm{d}r$

Pretesting for Size Distortions of the T-Test (Cont.)

- If $\delta = 0$ the first term of (3) vanishes and the distribution under the null is the usual $\mathrm{N}(0, 1)$.
- Likewise if $c << 0$, x is not persistent and first-order asymptotics are valid.

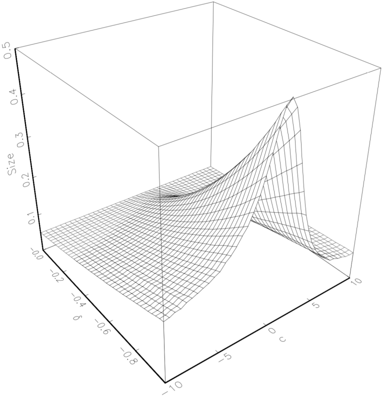


Figure 1: $p(c, \delta; 0.05) = \Pr\left(\delta \frac{\tau_c}{\kappa_c} + \left(1 - \delta^2\right)^{1/2} Z > z_{0.05}\right)$ Asymptotic Size of the One-Sided t-test at 5% Significance

Pretesting for Size Distortions of the T-Test (Cont.)

- CY propose that we may accept an actual size below 0.075.
- Test against the null of actual size greater than 0.075.
- Estimate δ from residuals of (1) and (2).
- Construct confidence interval for c by computing Dickey-Fuller generalized least squares (DF-GLS) test statistic[2] and using its known distribution under the alternative to construct the confidence interval for c.
- Reject the null if the confidence interval for c lies strictly below (or above) the region of the parameter space (c_{min}, c_{max}).
- Implemented in R package{pr}.

[2] For a detailed description of the DF-GLS statistic see the appendix

Developing a More Efficient Test

- To improve confidence of inference, we need to increase the signal-to-noise ratio. Here, we focus on reducing the noise.
- Since the innovations of (1) and (2) are correlated, CY propose to subtract off this part of the innovations.
- In contrast to the t-test, this procedure takes advantage of all the information contained in the system.
- Assume for the following $\alpha = \gamma = 0$ and ρ is known.

The Q-test

- Recall the joint log likelihood:

$$L(\beta,\rho,\alpha,\gamma) = -\frac{1}{1-\delta^2}\sum_{t=1}^{T}\Bigg[\frac{\left(r_t-\alpha-\beta x_{t-1}\right)^2}{\sigma_u^2} - 2\delta\frac{\left(r_t-\alpha-\beta x_{t-1}\right)\left(x_t-\gamma-\rho x_{t-1}\right)}{\sigma_u\sigma_e} + \frac{\left(x_t-\gamma-\rho x_{t-1}\right)^2}{\sigma_e^2}\Bigg]$$

- By the Neyman–Pearson Lemma the most powerful test against the simple alternative (i.e., a alternative that uniquely specifies the distribution) $\beta = \beta_1$ rejects the null if the LR is greater than some constant:

$$\sigma_u^2\left(1-\delta^2\right)\left(L\left(\beta_1\right)-L\left(\beta_0\right)\right) = 2\left(\beta_1-\beta_0\right)\sum_{t=1}^{T} x_{t-1}\left[r_t-\beta_{ue}\left(x_t-\rho x_{t-1}\right)\right] - \left(\beta_1^2-\beta_0^2\right)\sum_{t=1}^{T} x_{t-1}^2 > C \tag{6}$$

- Where $\beta_{ue} = \sigma_{ue}/\sigma_e^2$

The Q-test

- Observe that $\sum_{i=1}^{T} x_{t-1}^2$ does not depend on β.
- CY condition the test on that statistic in order to reduce it to

$$\sum_{t=1}^{T} x_{t-1}\left[r_t - \beta_{ue}\left(x_t - \rho x_{t-1}\right)\right] > C \tag{7}$$

- To get a standard normal distribution under the null they recenter and rescale:

$$\frac{\sum_{t=1}^{T} x_{t-1}\left[r_t - \beta_0 x_{t-1} - \beta_{ue}\left(x_t - \rho x_{t-1}\right)\right]}{\sigma_u\left(1-\delta^2\right)^{1/2}\left(\sum_{t=1}^{T} x_{t-1}^2\right)^{1/2}} > C \tag{8}$$

- Finally, after CY de-mean x_{t-1} and denote it by x_{t-1}^{μ} they can eliminate the assumption $\alpha = \gamma = 0$ to get:

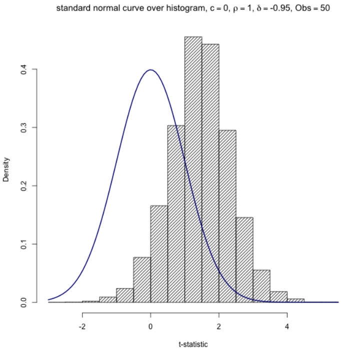
$$Q\left(\beta_0, \rho\right) = \frac{\sum_{t=1}^{T} x_{t-1}^{\mu}\left[r_t - \beta_0 x_{t-1} - \beta_{ue}\left(x_t - \rho x_{t-1}\right)\right]}{\sigma_u\left(1-\delta^2\right)^{1/2}\left(\sum_{t=1}^{T} x_{t-1}^{\mu 2}\right)^{1/2}} \tag{9}$$

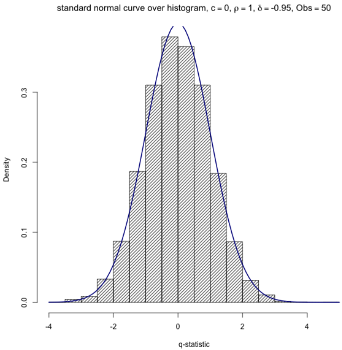
Intuition for Q-test

- Consider the case $\beta_0 = 0$.
- Then
$$Q(\beta_0, \rho) = \frac{\sum_{t=1}^{T} x_{t-1}^{\mu}\left[r_t - \beta_0 x_{t-1} - \beta_{ue}(x_t - \rho x_{t-1})\right]}{\sigma_u (1-\delta^2)^{1/2}\left(\sum_{t=1}^{T} x_{t-1}^{\mu 2}\right)^{1/2}}$$
is the t-statistic of the coefficient b regressing
$$r_t - \beta_{ue}(x_t - \rho x_{t-1}) = a + b x_{t-1} + v_t$$

- $x_t - \rho x_{t-1} = e_t + \gamma$
- $\beta_{ue} = \sigma_{ue}/\sigma_e^2$ tells us something about the relation between shocks of (1) and (2).
- The equation above can be interpreted as regressing the de-noised returns onto the regressor x where we exploit the information contained in ρ and the correlation of the shocks.

Making the Test Feasible





- If ρ and δ are known, this Q-test is the best we can do. It's UMP.
- In practice, however, ρ and δ are not known.
- ρ cannot be estimated consistently.
- CY use Bonferroni's inequality to get confidence intervals.

Bonferroni Confidence Intervals

- Construct a $(1-\alpha_1)$ confidence interval for ρ: $C_\rho(\alpha_1)$
- For each value of ρ in the confidence interval, construct a $(1-\alpha_2)$ confidence interval for β given ρ: $C_{\beta|\rho}(\alpha_2)$
- Taking the union over all $\rho \in C_\rho(\alpha_1)$ we can marginalize ρ:

$$C_\beta(\alpha) = \bigcup_{\rho \in C_\rho(\alpha_1)} C_{\beta|\rho}(\alpha_2) \tag{10}$$

- with $\alpha = \alpha_1 + \alpha_2$, $C_\beta(\alpha)$ has coverage of at least $(1-\alpha)$
- This follows from Bonferroni's inequality:

$$P\left(\bigcup_{i=1}^{n} E_i\right) \leq \sum_{i=1}^{n} P(E_i)$$

where $P(E_i)$ is the probability that E_i is true and $P\left(\bigcup_{i=1}^{n} E_i\right)$ is the probability that at least one of $E_1, E_2, \ldots, E_n$ is true [3].

---

[3] Weisstein, Eric W. "Bonferroni Inequalities." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/BonferroniInequalities.html

Bonferroni Confidence Intervals (Cont.)

- To get the confidence interval for ρ, CY need a unit root test statistic.
- Since they suspect ρ in the neighborhood of 1, the DF-GLS test statistic[4] is a good choice.
- To get the confidence interval for β they use the Q-test since they know it to be more powerful than the t-test given true ρ by the Neyman–Pearson Lemma. They hope it remains more powerful for other ρ as well. Whether this hope is met will be seen by nummerical analysis.

[4] For a detailed description of the DF-GLS statistic see the appendix

Bonferroni Confidence Intervals (Cont.)

$$C_{\beta|\rho}\left(\alpha_2\right)=\left[\underline{\beta}\left(\rho,\alpha_2\right),\overline{\beta}\left(\rho,\alpha_2\right)\right]$$

- where

$$\beta(\rho)=\frac{\sum_{t=1}^{T}x_{t-1}^{\mu}\left[r_t-\beta_{ue}\left(x_t-\rho x_{t-1}\right)\right]}{\sum_{t=1}^{T}x_{t-1}^{\mu 2}})$$

$$\underline{\beta}\left(\rho,\alpha_2\right)=\beta(\rho)-z_{\alpha_2/2}\sigma_u\left(\frac{1-\delta^2}{\sum_{t=1}^{T}x_{t-1}^{\mu 2}}\right)^{1/2}$$

$$\overline{\beta}\left(\rho,\alpha_2\right)=\beta(\rho)+z_{\alpha_2/2}\sigma_u\left(\frac{1-\delta^2}{\sum_{t=1}^{T}x_{t-1}^{\mu 2}}\right)^{1/2}$$

$z_{\alpha_2/2}$ denotes the $1-\alpha_2/2$ quantile of the standard normal distribution.

Bonferroni Confidence Intervals (Cont.)

- Let $C_{\rho}(\alpha_1) = \left[\underline{\rho}(\underline{\alpha}_1), \overline{\rho}(\overline{\alpha}_1)\right]$ denote the confidence interval for ρ,
- where $\underline{\alpha}_1 = \Pr\left(\rho < \underline{\rho}(\underline{\alpha}_1)\right)$, $\overline{\alpha}_1 = \Pr(\rho > \overline{\rho}(\overline{\alpha}_1))$, and $\alpha_1 = \underline{\alpha}_1 + \overline{\alpha}_1$.
- Then the Bonferroni confidence interval is:

$$C_{\beta}(\alpha) = \left[\underline{\beta}(\overline{\rho}(\overline{\alpha}_1), \alpha_2), \overline{\beta}\left(\underline{\rho}(\underline{\alpha}_1), \alpha_2\right)\right] \tag{11}$$

We Can Do Better

- The Bonferroni confidence interval is likely to be conservative, i.e., $\Pr(\beta \notin C_\beta(\alpha)) \leqslant \alpha_2(1-\alpha_1) + \alpha_1 \leqslant \alpha$ is likely a strict inequality.
- To see this consider:

$$\begin{aligned} \Pr(\beta \notin C_\beta(\alpha)) = & \Pr(\beta \notin C_\beta(\alpha) | \rho \in C_\rho(\alpha_1)) \Pr(\rho \in C_\rho(\alpha_1)) \\ & + \Pr(\beta \notin C_\beta(\alpha) | \rho \notin C_\rho(\alpha_1)) \Pr(\rho \notin C_\rho(\alpha_1)) \end{aligned} \tag{12}$$

- $\Pr(\beta \notin C_\beta(\alpha) | \rho \notin C_\rho(\alpha_1))$ is unknown. Thus we have to assume the worst case and bound it by one.
- $\Pr(\beta \notin C_\beta(\alpha) | \rho \in C_\rho(\alpha_1)) \leq \alpha_2$ is strict if $C_{\beta|\rho}(\alpha_2)$ depends on ρ.
- Bonferroni confidence interval is conservative since it is built on worst case scenario and reality is likely less harsh (more convervative the smaller is δ in absolute value).
- Refine confidence interval for ρ until the confidence interval for β is exactly at the desired significance level $\widetilde{\alpha}$.

Refined Bonferroni Method

- Do this by numerically searching over a grid.

1. fix α_2.
2. for each ρ, numerically search to find the $\overline{\alpha}_1$ s.t. :

- Pr $\left(\underline{\beta}\left(\overline{\rho}\left(\overline{\alpha}_1\right), \alpha_2\right) > \beta\right) \leq \widetilde{\alpha}/2$ holds for all values of c on the grid,
- and Pr $\left(\underline{\beta}\left(\overline{\rho}\left(\overline{\alpha}_1\right), \alpha_2\right) > \beta\right) = \widetilde{\alpha}/2$ at some point on the grid.

3. repeat 2. for $\underline{\alpha}_1$ s.t. Pr $\left(\overline{\beta}\left(\underline{\rho}\left(\underline{\alpha}_1\right), \alpha_2\right) < \beta\right) \leq \widetilde{\alpha}/2$

- $[\underline{\alpha}_1, \overline{\alpha}_1]$ is a tighter confidence interval for ρ.
- one-sided Bonferroni test[5] has exact size $\widetilde{\alpha}/2$ for some c.
- two-sided Bonferroni test has at most size $\widetilde{\alpha}$ for all c.

[5] For a detailed description of how to implement this test using OLS refer to the appendix. Alternatively, read the source code of R package{pr}

Comparing Power

- All tests considered should reject alternatives of the form $\beta = \beta_0 + b$, where b is some constant, almost surely as $T \to \infty$.
- More interesting are alternatives of the form $\beta = \beta_0 + b/T$, where b is again some constant.
- Under the local alternative

$$Q(\beta_0, \widetilde{\rho}) = \frac{b\left(T^{-2}\sum_{t=1}^{T} x_{t-1}^{\mu 2}\right)^{1/2}}{\sigma_u\left(1-\delta^2\right)^{1/2}} + \frac{\delta(\widetilde{c}-c)\left(T^{-2}\sum_{t=1}^{T} x_{t-1}^{\mu 2}\right)^{1/2}}{\omega\left(1-\delta^2\right)^{1/2}} + \frac{T^{-1}\sum_{t=1}^{T} x_{t-1}^{\mu}\left(u_t - \sigma_{ue}/\left(\sigma_e\omega\right)v_t\right) + \frac{1}{2}\sigma_{ue}/\left(\sigma_e\omega\right)\left(\omega^2-\sigma_v^2\right)}{\sigma_u\left(1-\delta^2\right)^{1/2}\left(T^{-2}\sum_{t=1}^{T} x_{t-1}^{\mu 2}\right)^{1/2}} \tag{13}$$

- where $\widetilde{c} = T(\widetilde{\rho}-1)$

Comparing Power (Cont.)

$$Q(\beta_0, \widetilde{\rho}) \Rightarrow \frac{b\omega\kappa_c}{\sigma_u (1-\delta^2)^{1/2}} + \frac{\delta(\widetilde{c}-c)\kappa_c}{(1-\delta^2)^{1/2}} + Z \tag{14}$$

- The power function for a right-tailed test is

$$\pi_Q(b) = \mathrm{E}\left[\Phi\left(z_\alpha - \frac{b\omega\kappa_c}{\sigma_u (1-\delta^2)^{1/2}} - \frac{\delta(\widetilde{c}-c)\kappa_c}{(1-\delta^2)^{1/2}}\right)\right] \tag{15}$$

- Where $\Phi(z)$ is one minus the standard normal CDF, z_α is the $1-\alpha$ quantile.
- The expectation is taken over the distribution of $(W_e(s), J_c(s))$.

Plotting the Power

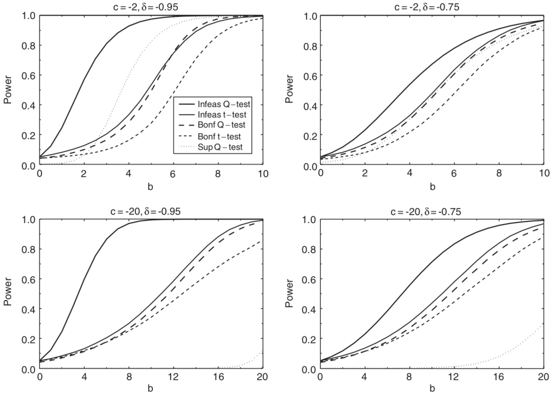


Figure 2: Null: $\beta = \beta_0$ against local alternatives: $b = T(\beta - \beta_0) > 0$ from Campbell and Yogo (2006)

- CY Compare:

1. A Bonferroni test based on the ADF test and the t-test.
2. A Bonferroni test based on the DF-GLS test and the t-test.
3. A Bonferroni test based on the DF-GLS test and the Q-test.

Analysis of the Power Gain (Cont.)

$c = -2$ and $\delta = -0.95$

- Pitman efficiency (relative number of observations needed to achieve 50% power) of test 1 relative to test 2 is 1.03.
- Pitman efficiency of test 2 relative to test 3 is 1.20.
- Close to no difference in power between the unrefined and refined Bonferroni t-test.
- Pitman efficiency of the unrefined relative to the refined Bonferroni Q-test is 1.62.

Analysis of the Power Gain (Cont.)

$c = -20$ and $\delta = -0.95$

- Pitman efficiency of test 1 relative to test 2 is 1.07.
- Pitman efficiency of test 2 relative to test 3 is 1.03.
- Pitman efficiency of the unrefined relative to the refined Bonferroni t-test is 1.23.
- Pitman efficiency of the unrefined relative to the refined Bonferroni Q-test is 1.55 .

Analysis of the Power Gain (Cont.)

- When the regressor is highly persistent, the use of the Q-test rather than the t-test is an important source of power gain for the Bonferroni Q-test.
- Bonferroni refinement is an especially important source of power gain for the Bonferroni Q-test since it tries to exploit information about ρ. This makes its confidence interval for β given ρ more sensitive to ρ. Hence, without the refinement the Bonferroni test is too conservative.

Finite Sample Rejection Rates (10,000 Monte Carlo runs)

Table 1: Finite-sample rejection rates for right-tailed tests of predictability at α=0.05

	Obs	c	ρ	δ	T-test	Bonf.Q-test	Q-test
1	50	0	1.000	-0.95	0.4160	0.0826	0.0483
2	50	0	1.000	-0.75	0.2916	0.0837	0.0515
3	50	-2	0.961	-0.95	0.2714	0.0868	0.0482
4	50	-2	0.961	-0.75	0.2079	0.0881	0.0532
5	50	-20	0.608	-0.95	0.0977	0.1206	0.0515
6	50	-20	0.608	-0.75	0.0840	0.1078	0.0484
7	100	0	1.000	-0.95	0.4217	0.0616	0.0480
8	100	0	1.000	-0.75	0.2930	0.0616	0.0497
9	100	-2	0.980	-0.95	0.2698	0.0587	0.0505
10	100	-2	0.980	-0.75	0.2104	0.0588	0.0489
11	100	-20	0.802	-0.95	0.1063	0.0622	0.0471
12	100	-20	0.802	-0.75	0.0874	0.0514	0.0500
13	250	0	1.000	-0.95	0.4259	0.0476	0.0483
14	250	0	1.000	-0.75	0.2970	0.0506	0.0536
15	250	-2	0.992	-0.95	0.2866	0.0507	0.0481
16	250	-2	0.992	-0.75	0.2092	0.0466	0.0492
17	250	-20	0.920	-0.95	0.1080	0.0406	0.0517
18	250	-20	0.920	-0.75	0.0944	0.0369	0.0501

Replicated and Updated Results

Table 2: Empirical results
Data was taken from Amit Goyal's Website. Stock returns are the SP 500 index log-returns from 1926 to 2017 from the Center for Research in Security Press (CRSP) minus the rolled over 3-month T-bill rate. ep is the log 10 year moving average earnings/price ratio (1926 to 2017). dp is the log dividend/price ratio (1926 to 2017). tbl is the 3-month T-bill rate (1952 to 2017). tms is the term-spread between long-term government bonds and tbl (1952 to 2017).

Prd	Regr	$\hat{\delta}$	CI $\hat{\rho}$	T-stat	Pt	$\hat{\beta}$	CI $\hat{\beta}$
Ann	ep	-0.97	[0.827,0.979]	2.12	0	0.114	[-0.01,0.18]
Ann	dp	-0.86	[0.875,0.986]	0.97	0	0.042	[-0.069,0.107]
Ann	tbl	0.08	[0.853,0.908]	-0.41	1	-0.279	[-0.125,0.075]
Ann	tms	-0.06	[0.454,0.575]	0.96	1	1.297	[-0.071,0.275]
Qua	ep	-0.98	[0.973,1.002]	3.16	0	0.048	[0.001,0.039]
Qua	dp	-0.95	[0.976,1.002]	1.82	0	0.023	[-0.012,0.026]
Qua	tbl	-0.09	[0.956,0.975]	-0.64	1	-0.099	[-0.045,0.019]
Qua	tms	0.06	[0.835,0.864]	1.44	1	0.489	[-0.005,0.104]
Mon	ep	-0.99	[0.993,1.002]	2.40	0	0.010	[-0.002,0.009]
Mon	dp	-0.98	[0.993,1.002]	1.20	0	0.004	[-0.005,0.006]
Mon	tbl	-0.13	[0.992,0.997]	-0.91	1	-0.043	[-0.013,0.003]
Mon	tms	0.04	[0.958,0.967]	1.49	1	0.156	[-0.001,0.032]

Results Reported by Campbell and Yogo (2006)

Tests of predictability

Series	Variable	t-stat	$\widehat{\beta}$	90% CI: β		Low CI β
				t-test	Q-test	($\rho = 1$)
Panel A: S&P 1880–2002, CRSP 1926–2002						
S&P 500	d–p	1.967	0.093	[−0.040, 0.136]	[−0.033, 0.114]	−0.017
	e–p	2.762	0.131	[−0.003, 0.189]	**[0.042, 0.224]**	−0.023
Annual	d–p	2.534	0.125	[−0.007, 0.178]	**[0.014, 0.188]**	0.020
	e–p	2.770	0.169	[−0.009, 0.240]	**[0.042, 0.277]**	0.002
Quarterly	d–p	2.060	0.034	[−0.014, 0.052]	[−0.009, 0.044]	−0.010
	e–p	2.908	0.049	[−0.001, 0.068]	**[0.010, 0.066]**	0.002
Monthly	d–p	1.706	0.009	[−0.006, 0.014]	[−0.005, 0.010]	−0.005
	e–p	2.662	0.014	[−0.001, 0.019]	**[0.002, 0.018]**	0.001
Panel C: CRSP 1952–2002						
Annual	d–p	2.289	0.124	[−0.023, 0.178]	[−0.007, 0.183]	0.020
	e–p	1.733	0.114	[−0.078, 0.178]	[−0.031, 0.229]	−0.025
	r_3	−1.143	−0.095	[−0.229, 0.045]	[−0.231, 0.042]	—
	y–r_1	1.124	0.136	[−0.087, 0.324]	[−0.075, 0.359]	−0.156
Quarterly	d–p	2.236	0.036	[−0.011, 0.051]	[−0.010, 0.030]	0.005
	e–p	1.777	0.029	[−0.019, 0.044]	[−0.012, 0.042]	−0.003
	r_3	−1.766	−0.042	**[−0.084, −0.004]**	**[−0.084, −0.004]**	−0.086
	y–r_1	1.991	0.090	**[0.009, 0.162]**	**[0.006, 0.158]**	−0.002
Monthly	d–p	2.259	0.012	[−0.004, 0.017]	[−0.004, 0.010]	0.001
	e–p	1.754	0.009	[−0.006, 0.014]	[−0.004, 0.012]	−0.001
	r_3	−2.431	−0.017	**[−0.030, −0.006]**	**[−0.030, −0.006]**	−0.030
	y–r_1	2.963	0.047	**[0.020, 0.072]**	**[0.020, 0.072]**	0.016

This table reports statistics used to infer the predictability of returns. Returns are for the annual S&P 500 index and the annual, quarterly, and monthly CRSP value-weighted index. The predictor variables are the log dividend–price ratio (d–p), the log earnings–price ratio (e–p), the three-month T-bill rate (r_3), and the long-short yield spread (y–r_1). The third and fourth columns report the t-statistic and the point estimate $\widehat{\beta}$ from an OLS regression of returns onto the predictor variable. The next two columns report the 90% Bonferroni confidence intervals for β using the t-test and Q-test, respectively. Confidence intervals that reject the null are in bold. The final column reports the lower bound of the confidence interval for β based on the Q-test at $\rho = 1$.

OOS results

Table 3: Empirical results OOS from 2003 to 2017.

Prd	Regr	$\hat{\delta}$	CI $\hat{\rho}$	T-stat	Pt	$\hat{\beta}$	CI $\hat{\beta}$
Qua	ep	-0.99	[0.79,1.015]	1.51	0	0.102	[-0.051,0.203]
Qua	dp	-0.97	[0.719,0.966]	0.56	0	0.037	[-0.097,0.198]
Qua	tbl	0.33	[0.95,0.987]	-0.70	1	-0.430	[-0.069,0.021]
Qua	tms	0.18	[0.9,0.949]	0.20	1	0.164	[-0.07,0.097]

R Implementation

- All source code is available at https://github.com/jpwoeltjen/PersistentRegressors
- Includes a R package that implements the methods discussed called pr (build it from source).
- Furthermore refer to the appendix for a mathematical implementation of the tests.

Appendix

DF-GLS

- Seeks power gain by assuming ρ is in the neighborhood of 1.
- Two possible alternative hypotheses: y_t is stationary around a linear trend or y_t is stationary with no linear time trend.
- Here we assume no linear time trend.
- Generate the following variables:
- $$\begin{aligned} \widetilde{y}_1 &= y_1 \\ \widetilde{y}_t &= y_t - \rho_{GLS} y_{t-1}, \quad & t = 2, \ldots, T \\ x_1 &= 1 \\ x_t &= 1 - \rho_{GLS}, & t = 2, \ldots, T \\ \rho_{GLS} &= 1 - 7/T \end{aligned}$$
- estimate by OLS $\widetilde{y}_t = \delta_0 x_t + \epsilon_t$
- The OLS estimator $\widehat{\delta}_0$ is then used to remove the mean from y_t that is, we generate

$$y^* = y_t - \widehat{\delta}_0$$

DF-GLS (Cont.)

- perform augmented Dickey–Fuller test on the transformed variable by fitting the OLS regression:

$$\Delta y_t^* = \beta y_{t-1}^* + \sum_{j=1}^{k-1} \zeta_j \Delta y_{t-j}^* + \epsilon_t$$

- For AR(1) this reduces to:

$$\Delta y_t^* = \beta y_{t-1}^* + \epsilon_t$$

- The t-statistic of β is the DF-GLS test statistic.
- Test the null hypothesis $H_0 : \beta = 0$ by using tabulated critical values.

Bonferroni Implementation $[\underline{\rho}, \overline{\rho}] = [1 + \underline{c}/T, 1 + \overline{c}/T]$

- Run the regression $r_t = \alpha + \beta x_{t-1} + u_t$ to get $\mathrm{SE}(\widehat{\beta})$.
- Run the regression $x_t = \gamma + \rho x_{t-1} + e_t$ to get $\mathrm{SE}(\widehat{\rho})$.
- Use the residuals $\widehat{u}_t$ and $\widehat{e}_t$ to compute:
- $$\widehat{\sigma}_u^2 = \frac{1}{T-2} \sum_{t=1}^{T} \widehat{u}_t^2$$
$$\widehat{\sigma}_e^2 = \frac{1}{T-2} \sum_{t=1}^{T} \hat{e}_t^2$$
$$\widehat{\sigma}_{ue} = \frac{1}{T-2} \sum_{t=1}^{T} \widehat{u}_t \widehat{e}_t$$
$$\widehat{\delta} = \frac{\widehat{\sigma}_{ue}}{\widehat{\sigma}_u \widehat{\sigma}_e}$$

Bonferroni Implementation $[\underline{\rho}, \overline{\rho}] = [1 + \underline{c}/T, 1 + \overline{c}/T]$ (Cont.)

- Compute the DF-GLS statistic.
- Given DF-GLS statistic and $\widehat{\delta}$ use lookup tables to get $[\underline{c}, \overline{c}]$.[6]
- Now we can compute the confidence interval for ρ which is given by $[\underline{\rho}, \overline{\rho}] = [1 + \underline{c}/T, 1 + \overline{c}/T]$

[6] Lookup tables are provided by Campbell, J.Y., Yogo, M., 2005. Implementing the econometric methods in "Efficient tests of stock return predictability". Unpublished working paper. University of Pennsylvania.

Bonferroni Implementation $[\underline{\beta}(\rho), \overline{\beta}(\rho)]$

- Run the regression $r_t^* = \alpha + \beta x_{t-1} + u_t$ for each $\rho = \{\underline{\rho}, \overline{\rho}\}$ to get $\widehat{\beta}(\rho)$.
- Where $r_t^* = r_t - \widehat{\sigma}_{ue}\widehat{\sigma}_e^{-2}(x_t - \rho x_{t-1})$.
- The confidence interval for β given ρ is $[\underline{\beta}(\rho), \overline{\beta}(\rho)]$
- Where

$$\underline{\beta}(\rho) = \widehat{\beta}(\rho) - 1.645\left(1 - \widehat{\delta}^2\right)^{1/2} \mathrm{SE}(\widehat{\beta})$$
$$\overline{\beta}(\rho) = \widehat{\beta}(\rho) + 1.645\left(1 - \widehat{\delta}^2\right)^{1/2} \mathrm{SE}(\widehat{\beta})$$

- The 90% Bonferroni confidence interval $[\underline{\beta}(\overline{\rho}), \overline{\beta}(\underline{\rho})]$ corresponds to a 10% two-sided test or a 5% one-sided test of the null hypothesis $\beta = 0$.