Empirical Asset Pricing: S01 Equity Return Predictability

Amedeo Andriollo

Finance Department (DRM)
Université Paris Dauphine - PSL.

Structure of the course

- Email: amedeo.andriollo@dauphine.psl.eu
- Grading follows a 30-70 rule: 70% final exam, 30% project/homework (3rd and 5th sessions).
- Dense slides. References at the end.
- Office hours: feel "free" to (DO) come.
- If you spot any typos/mistakes, please let me know: slides are updated regularly.
- This session:
  - about: 1) Warm up in Metrics and AP; 2) Campbell and Shiller (1988)'s decomposition; 3) Long-run regressions; 4) Cay.
  - built partially on: Cochrane (2008), Campbell (2014), and Pedersen's.

▷ More updates will follow.

A Quick Warm Up: Econometrics (1)

- Sample: the data we actually observe, as drawn from a population ("in the clouds").
  - We view sampled data as random, and the full population as fixed.
- Estimand: a function of distribution of observable data in the population.
  - A fixed number, again because the population is fixed
- Estimator: a function of data for the sample: what we actually see.
  - A random object, because the sampled data are random.
- Estimate: the value that the estimator takes when applied to a realized sample

$\hookrightarrow$ Estimand is the population parameter to measure VS. Estimator is the mathematical formula on sample/data ($\widehat{hat}$) VS. Estimate is the number after "applying" the estimator.

- (Unconditional) Moments:
  - Mean/Expectation: the probability-weighted value of X: $\mathbb{E}[X] = \sum_i x_i Pr(X = x_i)$
  - Variance: the squared spread of a distribution (from the mean): $Var[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$
  - Covariance: the linear association between two variables:

$$Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])], \quad Corr(X, Y) = Cov(X, Y)/(Std(X)Std(Y))$$

A Quick Warm Up: Econometrics (2)

- Quantiles: the $k-$th $q-$quantile is the data value where the cumulative distribution function crosses k/q: $Pr[X \leq x] \leq k/q$.

- Particular interest in econometrics and asset pricing is placed on the **conditional** expectation: the mean of the conditional distribution Y given X: $\mathbb{E}[Y|X = x]$, or $\mathbb{E}[Y|X]$.
  $\hookrightarrow$ Usually we denote: $\mathbb{E}_{t-1}[Y_t] = \mathbb{E}[Y_t|\mathcal{I}(t-1)]$

- Law of Iterated Expectations: $\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$.

- We say that Y is mean independent of X if $\mathbb{E}[Y|X] = \mathbb{E}[Y]$ (for time series: martingale property), and **uncorrelated** to X if $Cov(X,Y) = 0$. (**HW**: which one implies the other?)

- Key property of expectations is linearity: $\mathbb{E}[f(X) + g(X)Y|X] = f(X) + g(X)\mathbb{E}[Y|X]$.

A Quick Warm Up: Econometrics (3)

- When two variables are **jointly** normally distributed:

$$\mathbb{E}[Y|X] = \mathbb{E}[Y] + \frac{Cov(X,Y)}{Var(X)}(X - \mathbb{E}[X]) = \beta_0 + \beta_1 X$$

$\hookrightarrow$ conditional expectations are linear projections. If approx. normal, then least squares as estimator of the conditional mean: $(\widehat{\beta}_1, \widehat{\beta}_1) = \arg\min_{\beta_0,\beta_1} \mathbb{E}\big[(Y_i - \beta_0 - \beta_1 X_i)^2\big]$

- Multivariate: $\widehat{\beta} = (Z^\top Z)^{-1} Z^\top y$, with the regression $y = Z\beta + \epsilon$.
  The projection onto Z is: $P = Z(Z^\top Z)^{-1} Z^\top$. Notice: $y = Py + (I - P)y = Z\beta + \epsilon$.

- Unbiasedness: $\mathbb{E}[\widehat{\beta} - \beta] = 0$. A sufficient condition is: $\mathbb{E}[Z^\top \epsilon] = 0$, or stricter: $\mathbb{E}[\epsilon|Z] = 0$.
- **HW**: Refresh: consistency, efficiency, and BLUE. [EXTRA]

- Panel Data vs. Time Series: variation is spread across individuals (cross-sections) and time.
  $\hookrightarrow$ usually, in panel data the dynamics are "killed": weak/strict exogeneity.

A Quick Warm Up: Econometrics (4)

- Stationary process vs. non-stationary process: are the mean and variance changing over time?
- Univariate time series:
  - An ideal ARMA(p,q) is $y_t = c + \sum_{j=1}^{p} \rho_j \phi_{t-j} + \epsilon_t + \sum_{k=1}^{q} \theta_k \epsilon_{t-k}$, or $\Phi(L) y_t = c + \Theta(L)\epsilon_t$.
  - The White Noise (WN): $\epsilon_t \sim (0, \sigma^2)$. **HW:** Weak vs. Strong WN.
  - Characteristic polynomials are $\Phi(z) = 0$ and $\Theta(z) = 0$, we check if the roots are out-/inside: we have that roots larger than 1 $\iff$ eigenvalues smaller than 1.
  - AR(1) example: $y_t = c + \phi y_{t-1} + u_t$, its expectation is: $\mathbb{E}[y_t] = c/(1-\phi)$, and variance: Var$[y_t] = \sigma^2/(1-\phi^2)$. To get there, with $|\alpha| < 1$: $\sum_{i=0}^{\infty} a^i = 1/(1-a)$.
  - **HW:** Stability, Invertibility (AR representation), Causal (MA representation).
- Multivariate time series:
  - An ideal VARMA(p,q) is $Y_t = C + \sum_{j=1}^{p} A_j Y_{t-j} + \Theta(L) U_t$.
  - The characteristic polynomials are in determinant terms: $\det(I - \sum_{j=1}^{p} A_j z_j) = 0$.
  - Companion form, from VAR(p) to VAR(1): $\widetilde{A} = [A_1, A_2, ..., A_p; I, 0, ..., 0; 0, I, ..., 0; 0, ..., I, 0]$.
  - **HW:** Integration vs. Cointegration.

A Quick Warm Up: Econometrics (5)

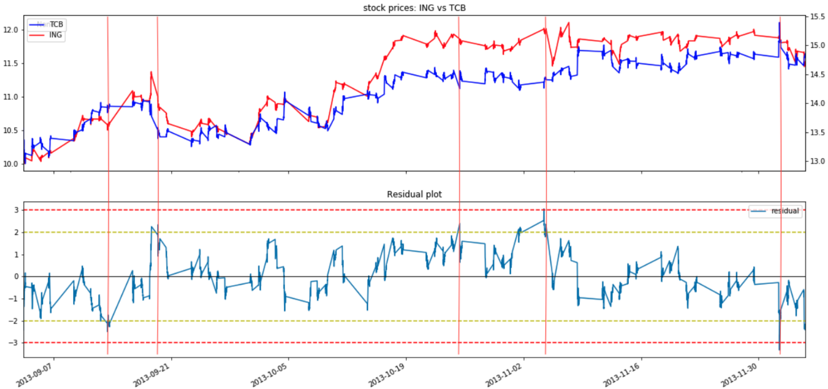
- Multivariate time series:
  - VAR(p) is $Y_t = C + \sum_{j=1}^{p} A_j Y_{t-j} + U_t$.
  - The characteristic polynomials are in determinant terms: $\det(I - \sum_{j=1}^{p} A_j z_j) = 0$.
- Non-stationarity. Integration and Cointegration:
  - Integration of order d: "how many (minimal) times we need to take differences to get a stationary series". d^{th} differencing: $(1-L)^d \implies I(d)$ series.

    $\hookrightarrow$ In the univariate case, we can use only the series itself.
  - Suppose we have two non-stationary $I(1)$ series: $\{x_t\}$ and $\{y_t\}$, and we fit a VAR(p).
    - A mechanical solution is to apply first difference to the univariate series and then fit the VAR.

    $\hookrightarrow$ However, if there exists a linear combination: $x_t + \beta y_t = u_t$, such that u_t is a stationary process, we do not need to transform the variables.
  - If cointegrated, the non-stationary series move together in the long run, and so remain bound to a stationary equilibrium (Engle and Granger, 1987).

  $\hookrightarrow$ No need to apply first difference to the univariate, go multivariate!

A Quick Warm Up: Econometrics (6)



- ING and TCB prices (non-stationary) move together, so their linear combination is stationary. [Source]

A Quick Warm Up: SDF (1)

- Any asset is defined by its state-contingent payoffs $\{X_s\}$, and state prices $\{q_s\}$ (=price of \$1 in a state s), for states $s = 1, ..., S$, in complete markets.
- **Law of one price** (No Arbitrage): two assets with identical payoffs in every state must have the same price. $\implies P = \sum_s q_s X_s$. Define the objective probabilities: $\{\pi_s\}$.
- Using the definition of the (**rational**) expectations, we can write:

$$P = \mathbb{E}[MX] = \sum_s \pi_s M_s X_s, \qquad M_s = q_s/\pi_s$$

  What is M? State Price Deflator, Pricing Kernel, or **Stochastic Discount Factor** (SDF).
- Remarks:
  1. A riskless asset pays 1 in every state $\implies P_f = \mathbb{E}[M] = 1/(1 + R_f)$.
  2. Theory? If you consider the classic two-period utility maximization problem of a price-taking investor, the F.O.C. (**Euler equation**):

$$u'(C_0)q_s = \beta\pi_s u'(C_s), \qquad s = 1, ..., S$$

A Quick Warm Up: SDF (2)

- For multiple assets in a multiperiod economy: $P_{i,t} = \mathbb{E}_t[M_{t+1} X_{i,t+1}]$
- Dividing by strictly positive prices, we have gross returns: $1 + R_{i,t+1} = X_{i,t+1}/P_{i,t}$
- We write:

$$\begin{aligned} 1 &= \mathbb{E}_t[M_{t+1}(1 + R_{i,t+1})] \\ &= \mathbb{E}_t[M_{t+1}]\mathbb{E}_t[1 + R_{i,t+1}] + \mathrm{Cov}_t(M_{t+1}, R_{i,t+1}) \\ &= (1 + R_{f,t+1})^{-1}\mathbb{E}_t[1 + R_{i,t+1}] + \mathrm{Cov}_t(M_{t+1}, R_{i,t+1}) \end{aligned}$$

and so: $\mathbb{E}_t[1 + R_{i,t+1}] = (1 + R_{f,t+1})(1 - \mathrm{Cov}_t(M_{t+1}, R_{i,t+1}))$, which means we can write:

$$1 + R_{i,t+1} = \mathbb{E}_t[1 + R_{i,t+1}] + U_{i,t+1}$$

where $\{U_{i,t+1}\}$ is a "noise": **unpredictable on avg** given the information at time t (and in particular, given the SDF).

Market Efficiency (1)

- "A market in which prices always 'fully reflect' available information is called *efficient*" (Fama, 1970)

$\hookrightarrow$ **Joint hypothesis** problem: market efficiency $\implies$ zero conditional mean for asset returns measured relative to the riskless interest rate and the **equilibrium** compensation for risk.

"..no other proposition in economics which has more solid evidence.." (Jensen, 1978).

- How to test it empirically: are returns (on avg) predictable?
- Towards Campbell and Shiller (1988), Shiller et al. (1981):

$$P_{i,t} = \mathbb{E}_t[M_{t+1}X_{i,t+1}] = \delta_t \mathbb{E}_t[X_{i,t+1}] = \delta_t \mathbb{E}_t[P_{i,t+1} + D_{i,t+1}] = \sum_{j=1}^{\infty} \mathbb{E}_t[\delta_t^j D_{i,t+j}]$$

where $\{D_{i,t}\}$ are the dividends. The **realized** discounted value of future dividends should equal the stock price plus unpredictable (uncorrelated) noise: $(\sum_{j=1}^{\infty} \delta_t^j D_{i,t+j}) = P_{i,t} + \epsilon_{t+1}$
$\hookrightarrow$ The realized discounted value of future dividends should have greater variance than the stock price ($\text{Var}(X+Y) \geq \text{Var}(X)$), but data tells the opposite!

Market Efficiency (2)

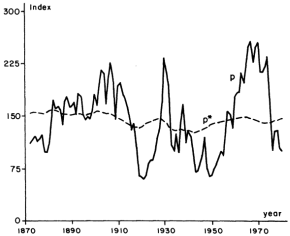


FIGURE 1
Note: Real Standard and Poor's Composite Stock Price Index (solid line p) and *ex post* rational price (dotted line p^*), 1871–1979, both detrended by dividing a long-run exponential growth factor. The variable p^* is the present value of actual subsequent real detrended dividends, subject to an assumption about the present value in 1979 of dividends thereafter. Data are from Data Set 1, Appendix.

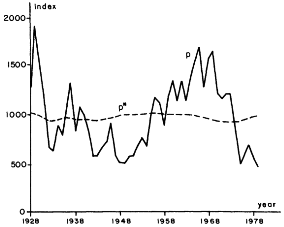


FIGURE 2
Note: Real modified Dow Jones Industrial Average (solid line p) and *ex post* rational price (dotted line p^*), 1928-1979, both detrended by dividing by a long-run exponential growth factor. The variable p^* is the present value of actual subsequent real detrended dividends, subject to an assumption about the present value in 1979 of dividends thereafter. Data are from Data Set 2, Appendix.

- *"The efficient markets model can be described as asserting that $p = \mathbb{E}[p^\star]$ [...] In other words, p, is the optimal forecast of $p^\star$. One can define the forecast error as $u_t = p_t^\star - p_t$ [...] If one uses the principle from elementary statistics that the variance of the sum of two uncorrelated variables is their sum of their variance [...] this means $var(p) \leq var(p^\star)$"* (Shiller et al., 1981).

Campbell and Shiller (1988)'s Decomposition (1)

- Connections between returns and dividends:

$$1 + R_{i,t+1} = (P_{i,t+1} + D_{i,t+1})/P_{i,t}, \quad R^e_{i,t+1} = (P_{i,t+1} + D_{i,t+1})/P_{i,t} - (1 + R_{f,t+1})$$

- Start with a (non-zero) return identity: $1 = (1 + R_{i,t+1})^{-1}(P_{i,t+1} + D_{i,t+1})P_{i,t}^{-1}$
- Multiply by the price-dividend ratio $P_{i,t}/D_{i,t}$:

$$\frac{P_{i,t}}{D_{i,t}} = (1 + R_{i,t+1})^{-1}\left(1 + \frac{P_{i,t+1}}{D_{i,t+1}}\right)\frac{D_{i,t+1}}{D_{i,t}}$$

- Take the logs: $p_{i,t} - d_{i,t} = -r_{i,t+1} + \Delta d_{i,t+1} + \ln(1 + e^{p_{i,t+1} - d_{i,t+1}})$
- Remarks on Logs:
  - Log returns? Note: for small enough $R_{i,t}$, we have: $r_{i,t} \approx R_{i,t}$. [Extra].
  - Note that I will start referring to log quantities, e.g. (log) price-dividend ratio.
  - Log price to dividend ratio is the negative of log dividend yield.

Campbell and Shiller (1988)'s Decomposition (2)

- Taylor approximations:
  - We have a nonlinear term: $\ln(1 + e^{p_{i,t+1} - d_{i,t+1}}) = f(p_{i,t+1} - d_{i,t+1})$
  - Taylor around the means: $f(x_t) \approx f(\mathbb{E}[x_t]) + (x_t - \mathbb{E}[x_t])\frac{d(f(x_t))}{d(x_t)}|_{x=\mathbb{E}[x_t]} + \cdots$
  - Objective:

$$f(p_{i,t+1} - d_{i,t+1}) \overset{Taylor}{\approx} \kappa_0 + \kappa_1(p_{i,t+1} - d_{i,t+1})$$

.

We have the following:

$$f(\mathbb{E}[x_t]) = \ln(1 + e^{\mathbb{E}[p_{i,t+1} - d_{i,t+1}]})$$

$$f'(x) = e^x/(1 + e^x) \implies \kappa_1 = e^{\mathbb{E}[p_{i,t+1} - d_{i,t+1}]}/(1 + e^{\mathbb{E}[p_{i,t+1} - d_{i,t+1}]})$$

↪ We conclude:

$$\ln(1 + e^{p_{i,t+1} - d_{i,t+1}}) \approx \overbrace{\ln(1 + e^{\mathbb{E}[p_{i,t+1} - d_{i,t+1}]}) - \kappa_1 \mathbb{E}[p_{i,t+1} - d_{i,t+1}]}^{\kappa_0} + \kappa_1(p_{i,t+1} - d_{i,t+1})$$

  - Avg. log price to dividend is around 3-3.5 (%) so $\kappa_1 \approx 0.95$: close to one.

Campbell and Shiller (1988)'s Decomposition (3)

- **This gives us the approximation**:

$$
\begin{aligned}
p_{i,t} - d_{i,t} &\approx \kappa_0 + \kappa_1 (p_{i,t+1} - d_{i,t+1}) - r_{i,t+1} + \Delta d_{i,t+1} \\
&\approx \kappa_0 + \kappa_1 \big(\kappa_0 + \kappa_1 (p_{i,t+2} - d_{i,t+2}) - r_{i,t+2} + \Delta d_{i,t+2}\big) - r_{i,t+1} + \Delta d_{i,t+1} \\
&\approx \kappa_0/(1-\kappa_1) - \sum_{j=0}^{\infty} \kappa_1^j r_{i,t+1+j} + \sum_{j=0}^{\infty} \kappa_1^j \Delta d_{i,t+1+j} + \lim_{j\to\infty} \kappa_1^j (p_{i,t+j+1} - d_{i,t+j+1})
\end{aligned} \tag{1}
$$

- To get there: rolling forward and transversality condition.
- Taking the conditional expectations:

$$
p_{i,t} - d_{i,t} = \frac{\kappa_0}{1-\kappa_1} - \underbrace{\sum_{j=0}^{\infty} \kappa_1^j \mathbb{E}_t[r_{i,t+1+j}]}_{\widetilde{r}_{i,t}} + \underbrace{\sum_{j=0}^{\infty} \kappa_1^j \mathbb{E}_t[\Delta d_{i,t+1+j}]}_{\widetilde{\Delta d}_{i,t}} .
$$

- $\widetilde{r}_{i,t}$: time-t discounted future returns of asset i.

Campbell and Shiller (1988)'s Decomposition (4)

- **Transversality condition**: $\lim_{j \to \infty} \kappa_1^j (p_{i,t+j+1} - d_{i,t+j+1}) = 0$
  $\hookrightarrow$ No explosive behavior of stock prices aka no bubbles! Giglio et al. (2016) use this condition to (model-free) test bubbles in the housing markets, using over 700 years.

- Discounted exp. returns as "discount-rate", and discounted exp. dividends as "cash-flow".

- We can say equivalently:
  1. Once current dividends ($d_{i,t}$) are realized, variation in the price-dividend ratio reflects news about either future dividend growth ($\Delta d_{i,t+1}$) or future returns ($r_{i,t+1}$).
  2. *"If we lived in an i.i.d. world, dividend yields would never vary in the first place"*. When they cannot move? If dividend growth and returns are not forecastable ("i.i.d. world"), aka $\mathbb{E}_t[\Delta d_{i,t+j}] = \mathbb{E}[\Delta d_{i,t}]$ and $\mathbb{E}_t[r_{i,t+j}] = \mathbb{E}[r_{i,t}] \implies p_{i,t} - d_{i,t}$ is constant, which is not(!)
     $\hookrightarrow$ *"Since dividend yields vary, they must forecast long-run returns, long-run dividend growth [...] Which of dividend growth or returns is forecastable?"* (Cochrane, 2008)

- Prices are **not** stationary, price-dividend ratios tend to be stationary.

Campbell and Shiller (1988)'s Decomposition (5)

- **HW**: substituting price, $p_{i,t}(d_{i,t}, \widetilde{r}_t, \widetilde{\Delta d}_t)$, in the approx.(1): only dividends and returns.

$$r_{i,t+1} - \mathbb{E}_t[r_{i,t+1}] = \sum_{j=0}^{\infty} \kappa_1^j \Big(\mathbb{E}_{t+1}[\Delta d_{i,t+1+j}] - \mathbb{E}_t[\Delta d_{i,t+1+j}]\Big)$$

$$- \sum_{j=1}^{\infty} \kappa_1^j \Big(\mathbb{E}_{t+1}[r_{i,t+1+j}] - \mathbb{E}_t[r_{i,t+1+j}]\Big)$$

$$= N_{i,t+1}^{CF}(\{\Delta d_{i,t+j}\}) + N_{i,t+1}^{DR}(\{r_{i,t+j}\})$$

where $N_{i,t+1}^{CF}$ denotes revisions in expectations about contemporaneous-to-$(t+1)$ and future discounted cash flows, $N_{i,t+1}^{DR}$ denotes revisions in expectations about future discount rates.

- What if today's returns ↑?
Suppose the AR(1): $r_{t+1} = \mathbb{E}_t[r_{t+1}] + u_{t+1} = \bar{r} + x_t + u_{t+1}, \quad x_{t+1} = \rho x_t + \epsilon_{t+1}$.
Then: $N_{i,t+1}^{DR} = \sum_{j=1} \kappa_1^j (\mathbb{E}_{t+1}[r_{t+1+j}] - \mathbb{E}_t[r_{t+1+j}]) = \epsilon_{t+1} \cdot \kappa_1 \cdot \sum_{j=0} \kappa_1^j \rho^j = \epsilon_{t+1} \dfrac{\kappa_1}{1 - \kappa_1 \rho}$

↪ If $\kappa_1 \approx 0.95$, then a 0.5% increase in the exp. return today (ϵ_{t+1}) is associated with a capital loss of about 0.67% when $\rho = 0.3$, a loss of 1.1% when $\rho = 0.6$.

Campbell and Shiller (1988)'s Decomposition (6)

- In a multivariate analysis (VAR), including the dividend-price ratio will make the same dynamic system invariant to whether you included return or dividend growth.
- Typical finding is that the volatility of N^{DR} is larger ($\approx \times 2$) than N^{CF}.
- As accounting identity, the approx.(1) holds ex-post $\implies$ it **should hold ex ante** as well, for any info set:

$$p_{i,t} - d_{i,t} = \frac{\kappa_0}{1-\kappa_1} + \mathbb{E}\left[\sum_{j=0}^{\infty} \kappa_1^j \Delta d_{i,t+1+j} - \sum_{j=0}^{\infty} \kappa_1^j r_{i,t+1+j} \middle| \mathcal{I}_t\right].$$

↪ Dividend yields ($=e^{d_{i,t}-p_{i,t}}$) reveal part of the **investor's information set**.
The price-dividend ratio measures the value, under rational expectations, of a very long-term (rational) investment strategy, aka the passive buy-and-hold.

The Variance of P/D Ratios (1)

- Let us calculate the variance of the price-dividend ratio:

$$\text{Var}(p_{i,t} - d_{i,t}) = \text{Var}(-\widetilde{r}_{i,t} + \widetilde{\Delta d}_{i,t}) = \text{Var}(\widetilde{r}_{i,t}) + \text{Var}(\widetilde{\Delta d}_{i,t}) - 2\text{Cov}(\widetilde{r}_{i,t}, \widetilde{\Delta d}_{i,t})$$

  If the price-dividend ratio is volatile, then either: 1) expected discounted future dividend growth is volatile, 2) expected discounted future returns are volatile, or 3) the two are negatively correlated. $\hookrightarrow$ **Excess volatility** puzzle (Shiller et al., 1981): prices (appear to) move more than what is implied by expected (or realized) dividends.

- Another way to get to the variance:

$$\text{Var}(p_{i,t} - d_{i,t}) = \text{Cov}(p_{i,t} - d_{i,t}, p_{i,t} - d_{i,t}) = \text{Cov}(p_{i,t} - d_{i,t}, -\widetilde{r}_{i,t} + \widetilde{\Delta d}_{i,t})$$
$$\implies 1 = \text{Cov}(p_{i,t} - d_{i,t}, \widetilde{\Delta d}_{i,t})/\text{Var}(p_{i,t} - d_{i,t}) - \text{Cov}(p_{i,t} - d_{i,t}, \widetilde{r}_{i,t})/\text{Var}(p_{i,t} - d_{i,t})$$

The Variance of P/D Ratios (2)

↪ Those are regression coefficients of long-run quantities!

1. Regress LR returns on price-dividend ratio: $\beta_r = \text{Cov}(\widetilde{r}_{i,t}, p_{i,t} - d_{i,t})/\text{Var}(p_{i,t} - d_{i,t})$.
2. Regress LR dividend growth on price-dividend ratio: $\beta_{\Delta d} = \text{Cov}(\widetilde{\Delta d}_{i,t}, p_{i,t} - d_{i,t})/\text{Var}(p_{i,t} - d_{i,t})$

- The volatility of price-dividend ratio thus corresponds to the ability to forecast returns and/or dividend growth (not just one year ahead but) many years ahead. In conclusion:
  1. β_r and $\beta_{\Delta d}$ measure how variation in the price-dividend ratio maps into variation in expected long-run returns (DR) and expected long-run dividend growth (CF).
  2. If we use the dividend yield $(d_{i,t} - p_{i,t})$ instead of the price-dividend ratio, the sign flips: the return-forecasting coefficient minus the dividend-growth-forecasting coefficient must be ≈ 1.

The Variance of P/D Ratios (3)

Table II
Long-Run Regression Coefficients

Table entries are long-run regression coefficients, for example, $b_r^{(k)}$ in $\sum_{j=1}^{k} \rho^{j-1} r_{t+j} = a + b_r^{(k)} dp_t + \varepsilon_{t+k}^r$. See equations (2)–(4). Annual CRSP data, 1947–2009. "Direct" regression estimates are calculated using 15-year ex post returns, dividend growth, and dividend yields as left-hand variables. The "VAR" estimates infer long-run coefficients from 1-year coefficients, using estimates in the right-hand panel of Table III. See the Appendix for details.

	Coefficient		
Method and Horizon	$b_r^{(k)}$	$b^{(k)}_{\Delta d}$	$\rho^k b^{(k)}_{dp}$
Direct regression , $k = 15$	1.01	−0.11	−0.11
Implied by VAR, $k = 15$	1.05	0.27	0.22
VAR, $k = \infty$	1.35	0.35	0.00

- From Table II: all price-dividend ratio **volatility corresponds to variation in expected returns**, not to dividend growth ("dog that does not bark") or bubbles (Cochrane, 2011).
- Cochrane's notation: $\kappa_1 = \rho$.

The Variance of P/D Ratios (4)

- What to take from Cochrane (2011)'s Table II?
  - Relative to today's dividends, high prices forecast *entirely* low future returns. (Cochrane, 2011)
  - This pattern of predictability is **pervasive across markets**:
    - [Bonds:] Variation in credit spreads across time/firms signals returns, not default probabilities.
    - [Houses:] High price-rent ratios signal low returns, not rising rents, (or not prices that rise forever = bubbles). [EXTRA]
  - Strong common/business cycle association:
    low (high) prices translate into high (low) expected returns, and this pattern is strongly related to business-cycles.

- High R^2 equals high predictability?[1] **Inference when overlapping** obs.:
  - Because of non-standard asymptotic properties: *"The tendency of long-run methods to produce "significant" results, no matter what the null hypothesis, should neither come as a surprise, nor be taken as conclusive evidence"* (Valkanov, 2003), thus **caution**...
  - ..vs. economic magnitude, Cochrane (2011): *"when point estimates are one and zero, arguing we should believe zero and one because zero and one cannot be rejected is a tough sell"*.

[1] $R^2 = 1 - \sum(y_t - \hat{y}_t)^2 / \sum(y_t - \bar{y})^2 = \widehat{\text{Corr}}(y,x), \quad \hat{y} = \hat{\beta}x, \quad R^2 = \beta^2(\sum x_t)/(\sum y_t)$

Bias when Regressors are Persistent (1)

- The previous regressions are **predictive** regression: $r_{t+1} = \alpha + \beta x_t + u_{t+1}$. Suppose that X is stationary AR(1): $x_t = \theta + \rho x_{t-1} + v_t$, with:

$$\text{Cov}\left(\begin{pmatrix} u_t \\ v_t \end{pmatrix}, (u_t, v_t)\right) = \Sigma = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \cdot & \sigma_v^2 \end{pmatrix}$$

- Recall that the estimator of β is unbiased when: $\mathbb{E}[\widehat{\beta} - \beta] = 0$.
- However, Stambaugh (1999) proved that, in finite-sample: (still **asymptotically** unbiased)

$$\mathbb{E}[\widehat{\beta} - \beta] = \frac{\sigma_{uv}}{\sigma_v^2}\mathbb{E}[\widehat{\rho} - \rho] = -\frac{\sigma_{uv}}{\sigma_v^2}\left[\frac{1 + 3\rho}{T}\right] + O(1/T^2)$$

  depending on the covariance between the return innovation (u_t) and the predictor innovation (v_t).
- Stambaugh (1999)'s Table 1: *"For the shortest and most recent period, still 20 years long ($T = 240$), $\widehat{\beta}$ has a bias (0.42) nearly as large as its standard deviation (0.45) and more than twice its realized value (0.19)"*.

Bias when Regressors are Persistent (2)

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 $N(0, \Sigma)$, 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

Bias when Regressors are Persistent (3)

- Lewellen (2004): worst-case bias-adjusted beta: $\widehat{\beta}_{adj} = \widehat{\beta} - \frac{\widehat{\sigma}_{uv}}{\widehat{\sigma}_v^2}(\widehat{\rho} - 1)$.
- Campbell and Yogo (2006): a possible solution is to add the regressor innovation:

$$r_{t+1} = \alpha + \beta x_t + \delta(x_{t+1} - \rho x_t) + v_{t+1}$$

↪ Why more efficiency? (**HW**: efficiency vs. consistency)

1. Consistent estimates: the extra regressor is uncorrelated with the original regressor, but...
2. ... it is **correlated with the future returns**: increased precision, as we control for some noise.

- Cochrane (2008): one "dog" has not "barked": $d_t - p_t$ **does not forecast** dividend growth:
  - Campbell and Shiller (1988)'s decomposition implies: $\beta_r = 1 + \beta_d - \rho\beta_{dp}$, (recall $\kappa_1 = \rho$) where the betas are regression of r_{t+1}, Δd_{t+1}, $d_{t+1} - p_{t+1}$ onto $d_t - p_t$.
  - If $\beta_{dp} < 1$ (dividend-price not explosive), the absence of dividend growth predictability can place a bound: say, $\rho = 0.96$, we have that $\beta_r = 0$ (unpredictable) implies: $\beta_d < -0.04$.
  - For simplicity, define the dividend yields: $dp_t = d_t - p_t$.

Bias when Regressors are Persistent (4)

- (..continuing) Suppose we have the following restricted VAR(1):

$$r_{t+1} = c_r + \beta_r dp_t + \epsilon^r_{t+1}, \quad \Delta d_{t+1} = c_d + \beta_d dp_t + \epsilon^d_{t+1}, \quad dp_{t+1} = c_{dp} + \beta_{dp} dp_t + \epsilon^{dp}_{t+1}$$

and if the variables are cointegrated, the Campbell and Shiller (1988)'s identity implies:

$$\underbrace{\beta_r dp_t + \epsilon^r_{t+1}}_{r_{t+1}} \approx -\kappa_1 \underbrace{\left(\beta_{dp} dp_t + \epsilon^{dp}_{t+1}\right)}_{dp_{t+1}} + dp_t + \underbrace{\left(\beta_d dp_t + \epsilon^d_{t+1}\right)}_{\Delta d_{t+1}}$$

therefore, matching coefficients and VAR residuals in the approximation gives two restrictions:

$$\beta_r = 1 + \beta_d - \kappa_1 \beta_{dp}, \quad \text{and} \quad \epsilon^r_{t+1} = -\kappa_1 \epsilon^{dp}_{t+1} + \epsilon^d_{t+1}$$

" Setting up a null in which returns are not forecastable, [...] I find that the absence of dividend-growth forecastability in our data provides much stronger evidence against this [...] Excess return forecastability is not a comforting result. [...] The only good piece of news is that observed return forecastability does seem to be just enough to account for the volatility of price dividend ratios. If both return and dividend-growth forecast coefficients were small, we would be forced to conclude that prices follow a "bubble" process, moving only on news (or, frankly, opinion) of their own future values" (Cochrane, 2008).

A Ca(y)se for Multivariate Prices (5)

Including more variables helps to have a broader info set (less small than investor's).

- Theory behind consumption/wealth ratio cay_t (Lettau and Ludvigson, 2001):
  - 2-period budget constraint: $W_{t+1} = (1 + R_{t+1})(W_t - C_t)$. Rolling it forwards and transversality condition: $W_t = C_t + \sum_{i=1} C_{t+i}/(\Pi_{j=1}^{i}(1 + R_{t+j}))$
  - Empirically? Loglinearize and the main ingredient is proxy for human capital: log wages, $\{y_t\}$ as the "dividend" paid on human capital investments. Together with the log asset holdings $\{a_t\}$:

$$c_t - \omega a_t - (1-\omega)y_t = \mathbb{E}_t\left[\sum_{i=1}^{\infty} \rho^i(\omega r_{t+1}^a + (1-\omega)r_{t+1}^h - \Delta c_{t+i})\right] + (1-\omega)z_t$$

  with ω being the avg share of asset holdings in total wealth.
  - The RHS variables are stationary $\implies$ cointegration of the LHS: common trend of c_t, a_t, y_t.
  - *"Investors who want to maintain a flat consumption path over time will attempt to "smooth out" transitory movements in their asset wealth arising from time variation in expected returns"* (Lettau and Ludvigson, 2001): good proxy for mkt expectations.

A Ca(y)se for Multivariate Prices (2)

Table IV

Forecasting Regressions with the Consumption-Wealth Ratio

Annual data 1952–2009. Long-run coefficients in the last two rows of the table are computed using a first-order VAR with dp_t and cay_t as state variables. Each regression includes a constant. *Cay* is rescaled so $\sigma(cay) = 1$. For reference, $\sigma(dp) = 0.42$.

	Coefficients		t-Statistics		Other Statistics		
Left-Hand Variable	dp_t	cay_t	dp_t	cay_t	R^2	$\sigma[E_t(y_{t+1})]\%$	$\frac{\sigma[E_t(y_{t+1})]}{E(y_{t+1})}$
r_{t+1}	0.12	0.071	(2.14)	(3.19)	0.26	8.99	0.91
Δd_{t+1}	0.024	0.025	(0.46)	(1.69)	0.05	2.80	0.12
dp_{t+1}	0.94	−0.047	(20.4)	(−3.05)	0.91		
cay_{t+1}	0.15	0.65	(0.63)	(5.95)	0.43		
$r_t^{lr} = \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}$	1.29	0.033				0.51	
$\Delta d_t^{lr} = \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j}$	0.29	0.033				0.12	

- What to take from Cochrane (2011)'s Table IV ?
  - Top: cay helps to forecast one-period returns but only marginally dividend growth.
  - Bottom: cay has almost no effect on LR quantities.
  - ↪ *"A variable can also help predict 1-year returns r_{t+1} without much changing long-run expected returns, if it has an offsetting effect on longer run returns $\{r_{t+j}\}$. Such a variable signals a change in the term structure of risk premia $\{\mathbb{E}_t[r_{t+j}]\}$"*(Cochrane, 2011).
  - cay forecasts +ive (-ive) near-term returns as it forecasts a long stream of -ive (+ive) LT returns.

Thanks for your attention! See you next Monday!

Bibliography I

Campbell, J. Y. (2014). Empirical asset pricing: Eugene fama, lars peter hansen, and robert shiller. *The Scandinavian Journal of Economics*, 116(3):593–634.

Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *The review of financial studies*, 1(3):195–228.

Campbell, J. Y. and Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of financial economics*, 81(1):27–60.

Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability. *The Review of Financial Studies*, 21(4):1533–1575.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of finance*, 66(4):1047–1108.

Engle, R. F. and Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, pages 251–276.

Fama, E. F. (1970). Efficient market hypothesis: A review of theory and empirical work. *Journal of Finance*, 25(2):28–30.

Giglio, S., Maggiori, M., and Stroebel, J. (2016). No-bubble condition: Model-free tests in housing markets. *Econometrica*, 84(3):1047–1091.

Jensen, M. C. (1978). Some anomalous evidence regarding market efficiency. *Journal of financial economics*, 6(2/3):95–101.

Lettau, M. and Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns. *the Journal of Finance*, 56(3):815–849.

Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics*, 74(2):209–235.

Shiller, R. J. et al. (1981). Do stock prices move too much to be justified by subsequent changes in dividends?

Stambaugh, R. F. (1999). Predictive regressions. *Journal of financial economics*, 54(3):375–421.

Valkanov, R. (2003). Long-horizon regressions: theoretical results and applications. *Journal of Financial economics*, 68(2):201–232.