Empirical Finance: Methods & Applications

Asset Return Predictability II

Pasquale Della Corte
Imperial College London

Week 4

Price-Dividend Relationship

Price-Dividend Relationship

Recall the gross return on an investment between times t and $t+1$ as

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t}$$

We can rearrange our identity as follows

$$P_t = \frac{P_{t+1} + D_{t+1}}{R_{t+1}}$$

Price-Dividend Relationship

Substituting for P_{t+1}, we obtain

$$P_t = \frac{D_{t+1}}{R_{t+1}} + \frac{P_{t+1}}{R_{t+1}}$$

$$= \frac{D_{t+1}}{R_{t+1}} + \frac{\frac{D_{t+2}}{R_{t+2}} + \frac{P_{t+2}}{R_{t+2}}}{R_{t+1}}$$

$$= \frac{D_{t+1}}{R_{t+1}} + \frac{D_{t+2}}{R_{t+1}R_{t+2}} + \frac{P_{t+2}}{R_{t+1}R_{t+2}}$$

Price-Dividend Relationship

Iterate forward

$$P_t = \frac{D_{t+1}}{R_{t+1}} + \frac{D_{t+2}}{R_{t+1}R_{t+2}} + \frac{D_{t+3}}{R_{t+1}R_{t+2}R_{t+3}} + \dots$$

while having that discounted value of P_{t+j}

$$\lim_{j\to\infty} \frac{P_{t+j}}{R_{t+1}R_{t+2}\dots R_{t+j}} = 0$$

goes to zero when j is very large.

Price-Dividend Relationship

We thus obtain that the price of an asset today is

$$P_t = \sum_{j=1}^{\infty} \frac{D_{t+j}}{\prod_{k=1}^{j} R_{t+k}}$$

simply the sum of the future discounted dividends.

More importantly, a high price at time t is associated with

- Low future returns and/or
- High future dividends.

Can we test it? Not really as prices are non stationary.

Price-Dividend Ratio

Divide every component by D_t and examine the price-dividend ratio as

$$\frac{P_t}{D_t} = \frac{1}{R_{t+1}} \frac{D_{t+1}}{D_t} + \frac{1}{R_{t+1} R_{t+2}} \frac{D_{t+2}}{D_t} + \frac{1}{R_{t+1} R_{t+2} R_{t+3}} \frac{D_{t+3}}{D_t} + \dots$$

which can further rewritten for convenience as

$$\frac{P_t}{D_t} = \frac{1}{R_{t+1}} \frac{D_{t+1}}{D_t} + \frac{1}{R_{t+1} R_{t+2}} \frac{D_{t+2}}{D_{t+1}} \frac{D_{t+1}}{D_t} + \frac{1}{R_{t+1} R_{t+2} R_{t+3}} \frac{D_{t+3}}{D_{t+2}} \frac{D_{t+2}}{D_{t+1}} \frac{D_{t+1}}{D_t} + \dots$$

Price-Dividend Ratio

In compact form, we have

$$\frac{P_t}{D_t} = \sum_{j=1}^{\infty} \prod_{k=1}^{j} \frac{D_{t+k} / D_{t+k-1}}{R_{t+k}}$$

The price-dividend ratio today is correlated

- positively with future dividend growth and/or
- negatively with future returns.

Can we test it? Not really as the relationship is non linear.

Log Price-Dividend Ratio

Campbell & Shiller (1988) propose a linear approximation of the price-dividend ratio

- It simply uses the first-order Taylor approximation.

Recall the gross return on an investment between times t and $t+1$ as

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t}$$

Multiply and divide by D_t and D_{t+1} as

$$P_t = \frac{P_{t+1} + D_{t+1}}{P_t} \times \frac{D_{t+1}}{D_{t+1}} \times \frac{D_t}{D_t}$$

Log Price-Dividend Ratio

Rearrange as

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{D_{t+1}} \times \frac{D_{t+1}}{D_t} \times \frac{D_t}{P_t},$$

and take the log transformation on both sides so that

$$\ln(R_{t+1}) = \ln\left(1 + \frac{P_{t+1}}{D_{t+1}}\right) + \ln\left(\frac{D_{t+1}}{D_t}\right) - \ln\left(\frac{P_t}{D_t}\right). \tag{1}$$

Note that I have inverted the last component in Equation (1) for convenience as

$$\ln\left(\frac{D_t}{P_t}\right) = -\ln\left(\frac{P_t}{D_t}\right).$$

Log Price-Dividend Ratio

1. The log of the simple gross return is the continuously compound return as

$$r_{t+1} = \ln(R_{t+1})$$

2. Rewrite the price-dividend ratio as

$$\frac{P_{t+1}}{D_{t+1}} = e^{\ln\left(\frac{P_{t+1}}{D_{t+1}}\right)} = e^{\ln(P_{t+1}) - \ln(D_{t+1})}$$

3. To further simplify the notation, use lowercase letter to denote log-variables as

$$x_t = \ln(X_t)$$

Log Price-Dividend Ratio

We can thus rewrite our identity in Equation (1) as

$$r_{t+1} = \ln\left[1 + e^{(p_{t+1} - d_{t+1})}\right] + (d_{t+1} - d_t) - (p_t - d_t)$$

Rename the log dividend growth as

$$\Delta d_{t+1} = d_{t+1} - d_t$$

Rename the log price-dividend ratio as

$$pd_{t+1} = p_{t+1} - d_{t+1}$$

Log Price-Dividend Ratio

We can thus rearrange our identity in Equation (1) as

$$r_{t+1} = \ln(1 + e^{pd_{t+1}}) + \Delta d_{t+1} - pd_t \tag{2}$$

The price-dividend relationship is nonlinear

- We use the first-order Taylor approximation to make it approximately linear,
- This approximation was initially proposed by Campbell & Shiller (1988),
- This procedure is also known as the Campbell-Shiller decomposition.

Log Price-Dividend Ratio

Consider the first-order Taylor approximation of $f(x)$ around a constant $\overline{x}$

$$f(x) \approx f(\overline{x}) + f'(\overline{x})(x - \overline{x})$$

In our case, we have that

$$\ln(1 + e^{x}) = \ln(1 + e^{\overline{x}}) + \frac{e^{\overline{x}}}{1 + e^{\overline{x}}}(x - \overline{x})$$

where

$$\overline{x} \longrightarrow \text{ long-run average of } x.$$

Log Price-Dividend Ratio

Replace x with pd_{t+1} and obtain

$$\ln(1 + e^{pd_{t+1}}) = \ln(1 + e^{\overline{pd}}) + \frac{e^{\overline{pd}}}{1 + e^{\overline{pd}}}(pd_{t+1} - \overline{pd})$$

where

$$\overline{pd} \longrightarrow \text{ long-run average of } pd_t.$$

Log Price-Dividend Ratio

Plug the linear approximation in Equation (2), and rewrite as

$$
\begin{aligned}
r_{t+1} &\approx \ln(1+e^{\overline{pd}}) + \frac{e^{\overline{pd}}}{1+e^{\overline{pd}}}(pd_{t+1} - \overline{pd}) + \Delta d_{t+1} - pd_t \\
&\approx \ln(1+e^{\overline{pd}}) - \frac{e^{\overline{pd}}}{1+e^{\overline{pd}}}\overline{pd} + \frac{e^{\overline{pd}}}{1+e^{\overline{pd}}} pd_{t+1} + \Delta d_{t+1} - pd_t .
\end{aligned}
$$

Log Price-Dividend Ratio

By setting

$$\kappa = \ln(1 + e^{\overline{pd}}) - \frac{e^{\overline{pd}}}{1 + e^{\overline{pd}}}\overline{pd}$$

and

$$\rho = \frac{e^{\overline{pd}}}{1 + e^{\overline{pd}}},$$

we can tidy up and obtain the Campbell-Shiller log-linear approximation

$$r_{t+1} \approx \kappa + \rho pd_{t+1} + \Delta d_{t+1} - pd_t.$$

What are the Implications for Predictability?

Rearrange as the (approximate) identity as

$$pd_t = \kappa + \rho pd_{t+1} + \Delta d_{t+1} - r_{t+1}.$$

Iterate forward and obtain the following present-value relationship

$$pd_t = \frac{\kappa}{1-\rho} + \sum_{j=0}^{\infty} \rho^j \Delta d_{t+j+1} - \sum_{j=0}^{\infty} \rho^j r_{t+j+1} \tag{3}$$

after imposing the no-Ponzi condition

$$\lim_{j\to\infty} \rho^j pd_{t+j} = 0$$

Using postwar US data, Campbell (1999) shows that ρ is about 0.964 in annual data.

What are the Implications for Predictability?

The present-value relationship must hold both ex-ante and ex-post (it arises from an identity)

$$pd_t = \frac{\kappa}{1-\rho} + E_t\left[\sum_{j=0}^{\infty} \rho^j \Delta d_{t+j+1}\right] - E_t\left[\sum_{j=0}^{\infty} \rho^j r_{t+j+1}\right] \tag{4}$$

This identity states that pd_t is high when

- Dividends are expected to grow rapidly in the future, and/or
- Stock returns are expected to be low in the future.

If the stock price is high today relative to its current dividend

- Investors must expect high dividends and/or low stock returns in the future.

What are the Implications for Predictability?

Multiply Equation (4) by $pd_t - E(pd_t)$ and take the expectations, giving

$$var(pd_t) = cov\left(pd_t, \sum_{j=0}^{\infty} \rho^j \Delta d_{t+j+1}\right) - cov\left(pd_t, \sum_{j=0}^{\infty} \rho^j r_{t+j+1}\right)$$

This equation states that all variation in the log price-dividend ratio must be explained by its covariance with future dividend growth and/or future returns

- The first *cov* is the slope of regressing future dividend growth rates on pd_t,
- The second *cov* is the slope of regressing future stock returns on pd_t.

What are the Implications for Predictability?

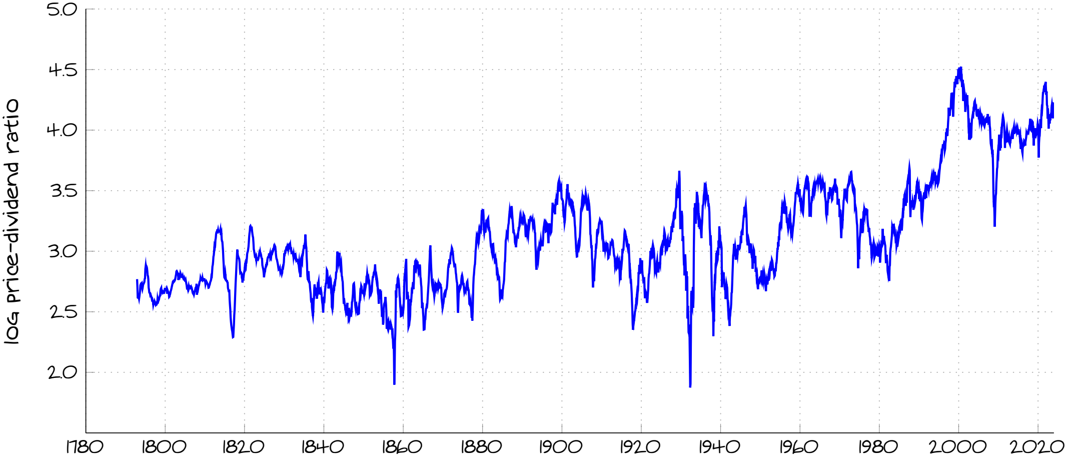
Decompose the variance of the log price-dividend ratio

$$var(pd_t) = cov\left(pd_t, \sum_{j=0}^{\infty} \rho^j \Delta d_{t+j+1}\right) - cov\left(pd_t, \sum_{j=0}^{\infty} \rho^j r_{t+j+1}\right)$$

If the log price-dividend ratio varies over time, then

- pd_t must positively predict future Δd_{t+j}
- pd_t must negatively predict future Δr_{t+j}

Log Price-Dividend Ratio



Data source: Global Financial Data.

What is the Empirical Evidence?

Table 1
Forecasting regressions

Regression	b	t	$R^2(\%)$	$\sigma(bx)(\%)$
$R_{t+1} = a + b(D_t/P_t) + \varepsilon_{t+1}$	3.39	2.28	5.8	4.9
$R_{t+1} - R_t^f = a + b(D_t/P_t) + \varepsilon_{t+1}$	3.83	2.61	7.4	5.6
$D_{t+1}/D_t = a + b(D_t/P_t) + \varepsilon_{t+1}$	0.07	0.06	0.0001	0.001
$r_{t+1} = a_r + b_r(d_t - p_t) + \varepsilon^r_{t+1}$	0.097	1.92	4.0	4.0
$\Delta d_{t+1} = a_d + b_d(d_t - p_t) + \varepsilon^{dp}_{t+1}$	0.008	0.18	0.00	0.003

R_{t+1} is the real return, deflated by the CPI, D_{t+1}/D_t is real dividend growth, and D_t/P_t is the dividend-price ratio of the CRSP value-weighted portfolio. R^f_{t+1} is the real return on 3-month Treasury-Bills. Small letters are logs of corresponding capital letters. Annual data, 1926–2004. $\sigma(bx)$ gives the standard deviation of the fitted value of the regression.

Cochrane (2008). "The Dog That Did Not Bark: A Defense of Return Predictability", *Review of Financial Studies*, 21, 1533-1575.

What is the Empirical Evidence?

Table I

Return-Forecasting Regressions

The regression equation is $R^e_{t\to t+k} = a + b \times D_t/P_t + \varepsilon_{t+k}$. The dependent variable $R^e_{t\to t+k}$ is the CRSP value-weighted return less the 3-month Treasury bill return. Data are annual, 1947–2009. The 5-year regression t-statistic uses the Hansen–Hodrick (1980) correction. $\sigma[E_t(R^e)]$ represents the standard deviation of the fitted value, $\sigma(\hat{b} \times D_t/P_t)$.

Horizon k	b	$t(b)$	R^2	$\sigma[E_t(R^e)]$	$\frac{\sigma[E_t(R^e)]}{E(R^e)}$
1 year	3.8	(2.6)	0.09	5.46	0.76
5 years	20.6	(3.4)	0.28	29.3	0.62

Cochrane (2011). "Presidential Address: Discount Rates", *Journal of Finance*, 66, 1047-1108.

What is the Empirical Evidence?

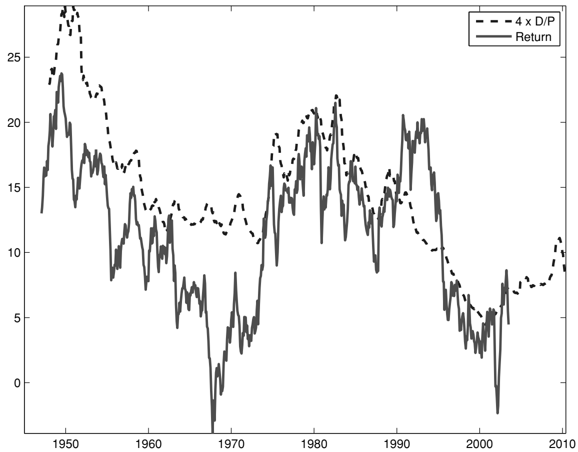


Figure 1. Dividend yield and following 7-year return. The dividend yield is multiplied by four. Both series use the CRSP value-weighted market index.

Cochrane (2011). "Presidential Address: Discount Rates", *Journal of Finance*, 66, 1047-1108.

Consumption-Wealth Ratio

Consumption-Wealth Ratio

Consider an economy where all wealth (including human capital) is tradable

- W_t is the aggregate wealth (human capital plus asset holdings),
- R_t is the gross return on the aggregate wealth,
- C_t is the aggregate consumption.

The period-by-period budget constraint an agent agent can be written as

$$W_{t+1} = R_{t+1}(W_t - C_t)$$

Consumption-Wealth Ratio

Solve forward with an infinite horizon and obtain

$$W_t = C_t + \sum_{i=1}^{\infty} \frac{C_{t+i}}{\prod_{j=1}^{i} R_{t+j}}$$

imposing the transversality condition that the limit of discounted future wealth is zero.

This equation says that today's wealth equals the discounted value of all future consumption

- The consumption-wealth relationship is nonlinear,
- Campbell and Mankiw (1989) propose a log-linear approximation.

Consumption-Wealth Ratio

Divide the budget constraint by W_t

$$\frac{W_{t+1}}{W_t} = R_{t+1}\left(1 - \frac{C_t}{W_t}\right),$$

and then take logs

$$w_{t+1} - w_t = r_{t+1} + \ln\left(1 - e^{c_t - w_t}\right)$$

while using lowercase letter to denote log variables.

Consumption-Wealth Ratio

Take a first-order Taylor expansion of the nonlinear term around $\overline{c_t - w_t}$ and obtain

$$\ln\left(1 - e^{c_t - w_t}\right) \approx \kappa + \left(1 - \frac{1}{\rho}\right)(c_t - w_t)$$

where

$$\kappa = \ln(1 - e^{\overline{c_t - w_t}}) - \left(1 - \frac{1}{\rho}\right)\ln(1 - \rho)$$

and

$$\rho = 1 - e^{\overline{c_t - w_t}}$$

The term ρ can be seen as wealth W as

$$\rho = 1 - \frac{\overline{C}}{\overline{W}} \longrightarrow \frac{\overline{W} - \overline{C}}{\overline{W}} < 1$$

Consumption-Wealth Ratio

We can thus rewrite the log budget constraint of the representative agent as

$$w_{t+1} - w_t = \kappa + \left(1 - \frac{1}{\rho}\right)(c_t - w_t) + r_{t+1}$$

By solving this difference equation, we can express the log consumption-wealth ratio as

$$c_t - w_t = \sum_{i=1}^{\infty} \rho^i (r_{t+i} - \Delta c_{t+i})$$

after imposing the transversality condition

$$\lim_{i \to \infty} \rho^i (c_{t+i} - w_{t+i}) = 0$$

Consumption-Wealth Ratio

The log consumption-wealth ratio must holds both ex-ante and ex-post

$$c_t - w_t = E_t \left[\sum_{i=1}^{\infty} \rho^i (r_{t+i} - \Delta c_{t+i}) \right]$$

A high log consumption-wealth ratio today must be associated with

- high future rates of return on invested wealth, and/or
- low future consumption growth.

The consumptionwealth ratio can only vary if consumption growth and/or returns are predictable.

Consumption-Wealth Ratio

As aggregate wealth is not observable, Lettau and Ludvigson (2001) propose to measure

- Aggregate wealth using asset holdings A_t and human capital H_t,
- Return on wealth using returns on asset holdings $R_{a,t}$ and labour income $R_{h,t}$

The log consumption-wealth ratio can we rewritten as

$$c_t - \omega a_t - (1-\omega) h_t = E_t \left[\sum_{i=1}^{\infty} \rho^i (\omega r_{a,t+i} + (1-\omega) r_{h,t+i} - \Delta c_{t+i}) \right]$$

Consumption-Wealth Ratio

Human capital H_t is not observable but we can use

$$h_t = \kappa + y_t + z_t$$

where the log of human capital h_t is related to the log of labour income y_t.

The log consumption-wealth ratio can we rewritten as

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

where the combined terms on both sides must be stationary.

Consumption-Wealth Ratio

The log consumption-wealth ratio is a present-value relationship

- It is subject to transversality condition,
- The combined terms on both sides must be stationary.

The individual term c_t, a_t, and y_t are nonstationary but their combination

$$c_t - \omega a_t - (1-\omega) y_t \quad \Longrightarrow \quad \text{must be stationary}$$

meaning that

$$c_t, \quad a_t, \text{ and } \quad y_t \quad \Longrightarrow \quad \text{must be cointegrated}$$

Consumption-Wealth Ratio

The deviation from the common trend is then called

$$cay_t = c_t - \omega a_t - (1 - \omega) y_t$$

cay_t is a proxy for market expectations of future asset returns $r_{a,t+i}$ as long as expected future returns on human capital $r_{h,t+i}$ and expected future consumption growth Δc_{t+i} are not too variable.

How to Estimate cay?

Lettau and Ludvigson (2001) use a dynamic ordinary least squares (DOLS)

$$c_t = \alpha + \beta_a a_t + \beta_y y_t + \sum_{i=-k}^{k} b_{a,i} \Delta a_{t-i} + \sum_{i=-k}^{k} b_{y,i} \Delta y_{t-i} + \varepsilon_t$$

which adds leads and lags of the first difference of the right-hand side variables to eliminate the effects of regressor endogeneity on the distribution of the least squares estimator.

We can then obtain

$$cay_t = c_t - \widehat{\beta}_a a_t + \widehat{\beta}_y y_t$$

How to Estimate cay?

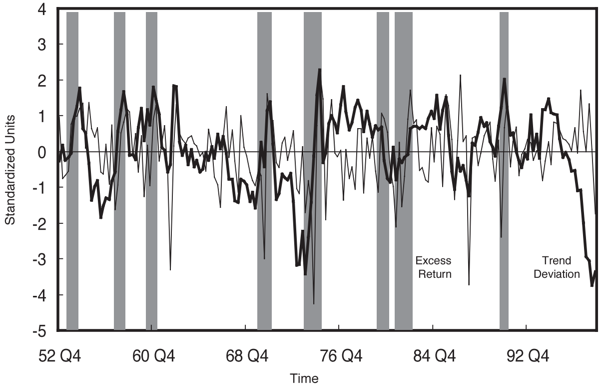


Figure 1. Excess returns and trend deviations. Excess return is the return on the S&P Composite Index less the return on the three-month Treasury bill rate. Trend deviation is the estimated deviation from the shared trend in consumption c, labor income y, and asset wealth a: $\widehat{cay}_t = c_t - \hat{\beta}_a a_t - \hat{\beta}_y y_t$. Both series are normalized to standard deviations of unity. The sample period is fourth quarter of 1952 to third quarter of 1998. Shaded areas denote NBER recessions.

Lettau and Ludvigson (2001). "Consumption, Aggregate Wealth, and Expected Stock Returns", *Journal of Finance*, 56, 815-849.

Does cay Predict Future Returns?

#	Constant (t-stat)	lag (t-stat)	$\widehat{cay}_t$ (t-stat)	$d_t - p_t$ (t-stat)	$d_t - e_t$ (t-stat)	$RREL_t$ (t-stat)	TRM_t (t-stat)	DEF_t (t-stat)	$\bar{R}^2$
				Panel A: Real Returns; 1952:4–1998:3					
1	**0.017**	**0.136**							0.01
	(3.131)	(2.221)							
2	**0.029**		**2.220**						0.09
	(4.672)		(3.024)						
3	**0.026**	0.062	**2.109**						0.09
	(4.645)	(0.981)	(2.806)						
4†	**0.028**	−0.007	**2.513**						0.10
	(4.889)	(−0.157)	(4.754)						
				Panel B: Excess Returns; 1952:4–1998:3					
5	**0.014**	**0.119**							0.00
	(2.952)	(1.976)							
6	**0.024**		**2.165**						0.09
	(4.328)		(3.226)						
7	**0.023**	0.043	**2.089**						0.09
	(4.345)	(0.707)	(2.988)						
8†	**0.022**	−0.038	**2.528**						0.10
	(4.612)	(−0.483)	(4.583)						

Lettau and Ludvigson (2001). “Consumption, Aggregate Wealth, and Expected Stock Returns“, *Journal of Finance*, 56, 815-849.

Out-of-Sample Evidence

tay versus cay

Brennan and Xia (2005) argued

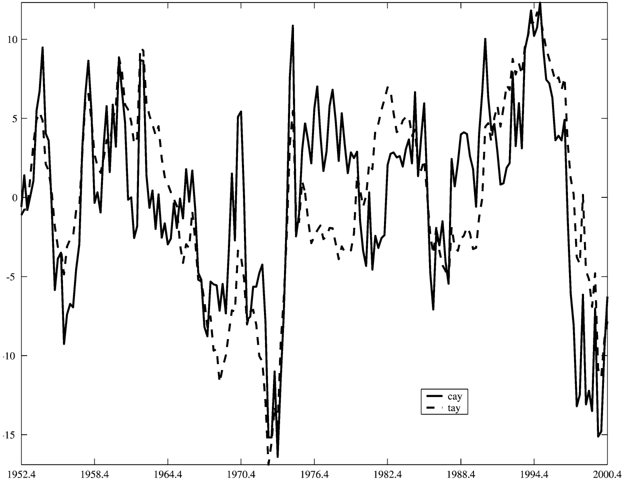
- The predictive power of cay_t arises from a "look-ahead bias",
- This happens as the parameters β_a and β_y are fitted in-sample.

They run a similar DOLS regression but replace consumption c_t with a linear trend t and obtain

$$tay_t = t - \widehat{\beta}_a a_t + \widehat{\beta}_y y_t$$

- tay_t forecast stocks returns as well as cay_t under "look-ahead bias",
- Both cay_t and tay_t lose their out-of-sample forecasting power when they are re-estimated every period with only available data.

tay versus cay



Brennan and Xia (2005). ‘tays as good as cay“, *Finance Research Letters*, 2, 1–14.

In-Sample Analysis

A. S&P real return 1952.4 to 2000.4

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
constant	0.010	0.011	−1.138	−0.882	−0.383	−0.053	0.010	0.011	0.011	0.011
	(1.98)	(2.11)	(3.19)	(2.41)	(0.75)	(0.11)	(1.73)	(2.02)	(1.76)	(1.96)
$\widehat{cay}_{t-1}$			1.874		0.642					
			(3.24)		(0.77)					
$\widehat{cay}_{t-2}$				1.457		0.103				
				(2.46)		(0.14)				
$\widehat{tay}_{t-1}$	0.004				0.003					
	(4.78)				(2.38)					
$\widehat{tay}_{t-2}$		0.004				0.004				
		(4.22)				(2.80)				
$\widehat{ta}_{t-1}$							0.001			
							(2.39)			
$\widehat{ca}_{t-1}$									0.026	
									(0.27)	
$\widehat{ty}_{t-1}$								0.002		
								(3.40)		
$\widehat{cy}_{t-1}$										0.600
										(2.30)
$\bar{R}^2$	0.100	0.077	0.076	0.043	0.099	0.072	0.025	0.051	−0.005	0.031

Brennan and Xia (2005). "tays as good as cay", *Finance Research Letters*, 2, 1–14.

Out-of-Sample Analysis

	Constant	$\widehat{cay}^{DLS}_{t-1}$	$\widehat{cay}^{OLS}_{t-1}$	$\widehat{tay}_{t-1}$	$\widehat{cay}^{DLS}_{t-2}$	$\widehat{cay}^{OLS}_{t-2}$	$\widehat{tay}_{t-2}$
	Root mean square error						
S&P real return	0.0837	0.0872	0.0846	0.0851	0.0868	0.0845	0.0840
S&P excess return	0.0817	0.0850	0.0828	0.0831	0.0845	0.0827	0.0822
		Pseudo R^2 (%)					
S&P real return		−8.46	−2.11	−3.25	−7.39	−1.94	−0.72
S&P excess return		−8.31	−2.62	−3.53	−7.08	−2.44	−1.27

Brennan and Xia (2005). "tays as good as cay", *Finance Research Letters*, 2, 1–14.

Out-of-Sample Evidence

Weltch and Goyal (2008) study the performance of several predictors for the equity premium,

- These predictors have predicted poorly both in-sample (IS) and out-of-sample (OOS),
- They are unstable, as diagnosed by their out-of-sample predictions and other statistics,
- They would have not have helped an investor with real-time information to earn a profit.

Unobserved Factors

Bayesian Methods

Bayesian econometrics is based on simple probability rules to

- Estimate the parameters of a model,
- Obtain the predictions from a model,
- Compare different models.

Key aspects of Bayesian econometrics

- Learning about something unknown (parameters) given something that is known (data),
- The *conditional probability* of parameters given data is the best way of summarizing what we have learned,

Bayesian Methods

The rule of probability for two random variables A and B

$$p(A, B) = p(A|B)\, p(B)$$

We can reverse the roles of A and B and obtain

$$p(A, B) = p(B|A)\, p(A)$$

- $p(A, B) \longrightarrow$ *joint probability* of A and B,
- $p(A|B) \longrightarrow$ *conditional probability* of A given B,
- $p(B|A) \longrightarrow$ *conditional probability* of B given A,
- $p(A) \longrightarrow$ *marginal probability* of A,
- $p(B) \longrightarrow$ *marginal probability* of B.

Bayesian Methods

We can the combine these expressions

$$p(A, B) = p(A|B)\, p(B)$$

and

$$p(A, B) = p(B|A)\, p(A)$$

We thus obtain the **Bayes rule**, which summarizes B given A as

$$p(B|A) = \frac{p(A|B)\, p(B)}{p(A)}$$

Bayesian Methods

A typical exercise in empirical applications

- y denotes the observed data,
- θ refers to the parameters of a model,
- We use y to learn about θ.

Bayesian econometrics uses the **Bayes rule** to obtain

$$p(\theta|y) = \frac{p(y|\theta)\, p(\theta)}{p(y)}$$

- $p(\theta|y)$ is equivalent to asking "given data y, what do we learn about the parameter θ?"
- θ is a random variable for *Bayesian econometrics* and a fixed point.

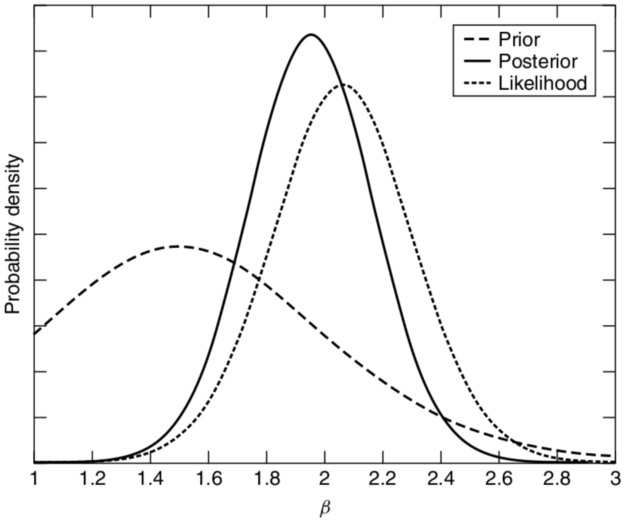
Bayesian Methods

As we only care about θ, we can ignore $p(y)$ (not involving θ) and rewrite the **Bayes rule** as

$$p(\theta|y) \propto p(y|\theta)\, p(\theta)$$

- $p(\theta|y) \longrightarrow$ posterior density of θ (what we have learned about an unknown quantity posterior to seeing the data),
- $p(y|\theta) \longrightarrow$ likelihood function (the density of the data given the parameters),
- $p(\theta) \longrightarrow$ prior density of θ (what the researcher knows about θ before observing the data),
- $p(y) \longrightarrow$ marginal density of y (the density of the data while ignoring information about θ),
- The symbol $\propto$ means proportional to.

Bayesian Methods



Bayesian Methods

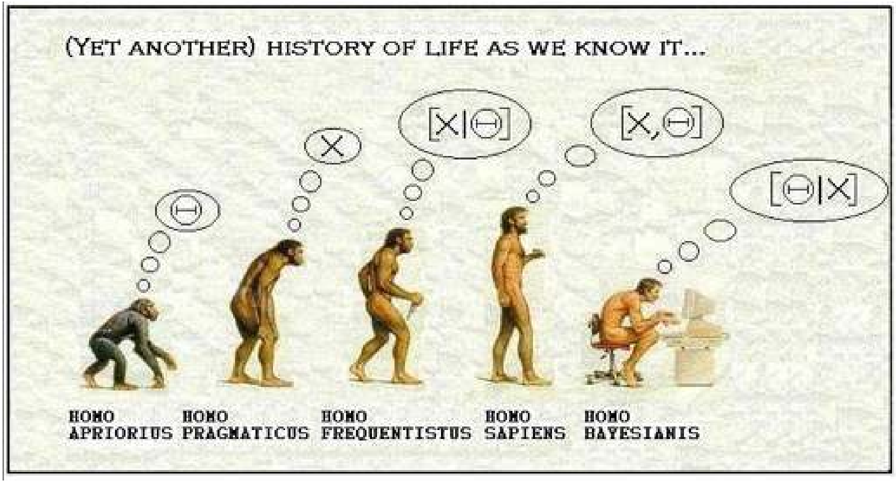
Bayes rule simply states that

$$p\left(\theta|y\right) \propto p\left(y|\theta\right) p\left(\theta\right)$$

"the posterior is proportional to the likelihood times the prior".

We can think of it as an updating rule, where the data allows us to update our prior views about θ. The result is the posterior which combines prior beliefs with the *current information* from the data.

Bayesian Methods



Introduction

The driving forces of economic and financial variables are often unobservable

- Asset prices may be driven by state variables (e.g., liquidity, volatility, and risk aversion) not directly measurable,
- Economic theory suggests that macro variables (e.g., economic growth) are driven by unobservable factors (e.g., technological change or human capital accumulation).

A state space is used to assess specifications with unobservable explanatory variables

- A system of equations with y_t (observable) and x_t (unobservable),
- The system of equations may be nonlinear and/or non-Gaussian.

State Space Model

A state space model consists of

- a measurement equation which links actual observations to (latent or unobserved) state variables
- a transition (or state) equation which describes the evolution of the state variable according to Markov process

A state space model allows

- an observed process y_t as being explained by unobserved state variables, which are driven by a stochastic process.

The Kalman Filter (Kalman, 1960)

- a particular algorithm that is used to solve state space models in a linear and Gaussian case
- we will derive the Kalman Filter using a Bayesian Approach.

Kalman Filter: Preliminary

Let X be normally distributed and consider its partition into X_1 and X_2 as

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m_1 \\ m_2 \end{bmatrix}, \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}\right).$$

The conditional distributions of X_1 given X_2 is

$$X_1 | X_2 \sim \mathcal{N}\left(m_1\left(X_2\right), V_{11}\left(X_2\right)\right)$$

where

$$m_1\left(X_2\right) = m_1 + V_{12} V_{22}^{-1}\left(X_2 - m_2\right)$$

and

$$V_{11}\left(X_2\right) = V_{11} - V_{12} V_{22}^{-1} V_{21}.$$

Kalman Filter: Preliminary

The conditional distributions of X_2 given X_1 is

$$X_2 | X_1 \sim \mathcal{N}\left(m_2\left(X_1\right), V_{22}\left(X_1\right)\right)$$

where

$$m_2\left(X_1\right) = m_2 + V_{21} V_{11}^{-1}\left(X_1 - m_1\right)$$

and

$$V_{22}\left(X_1\right) = V_{22} - V_{21} V_{11}^{-1} V_{12}.$$

Kalman Filter: An Example

The measurement and state equations are defined as

$$r_t = Z_{t-1}\mu_{t-1} + X_{t-1}\beta + \varepsilon_t$$

$$\mu_t = c + \Phi\mu_{t-1} + u_t$$

- $r_t \longrightarrow 1 \times 1$ return at time t,
- $Z_t \longrightarrow 1 \times K$ vector of regressors at time t,
- $\mu_t \longrightarrow K \times 1$ vector of unobserved component,
- $X_t \longrightarrow 1 \times M$ vector of regressors at time t,
- $\beta \longrightarrow M \times 1$ vector of constant slopes,
- $c \longrightarrow K \times 1$ vector of parameters.
- $\Phi \longrightarrow K \times K$ matrix of parameters (with eigenvalues $|\lambda_i| < 1$ for $i = 1, \ldots, K$).

Kalman Filter: An Example

The errors are normally distributed and serially uncorrelated

$$\begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} & 0 \\ 0 & \Sigma_{uu} \end{bmatrix} \right).$$

This example nests two special cases:

1. Regression with time-varying parameters,
2. Regression with an unobserved predictor.

Regression with Time-Varying Parameters

Suppose you interested in time-varying slope coefficients γ and δ

$$r_t = \alpha + \gamma_{t-1} dp_{t-1} + \delta_{t-1} cay_{t-1} + \varepsilon_t$$

$$\gamma_t = c_1 + \phi_1 \gamma_{t-1} + u_{1,t}$$

$$\delta_t = c_2 + \phi_2 \delta_{t-1} + u_{2,t}$$

which is equivalent to

$$Z_{t-1} = \begin{bmatrix} dp_{t-1} & cay_{t-1} \end{bmatrix}, \quad X_{t-1} = 1 \quad \beta = \alpha$$

$$\mu_t = \begin{bmatrix} \gamma_t \\ \delta_t \end{bmatrix}, \quad \Phi = \begin{bmatrix} \phi_1 & 0 \\ 0 & \phi_2 \end{bmatrix}, \quad u_t = \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}$$

Regression with an Unobserved Predictor

Suppose you interested in an unobserved predictor μ

$$r_t = \alpha + \gamma dp_{t-1} + \mu_{t-1} + \varepsilon_t$$

$$\mu_t = \phi \pi_{t-1} + u_t$$

which is equivalent to

$$X_{t-1} = [1 \quad dp_{t-1}], \quad Z_{t-1} = 1, \quad \Phi = \phi$$

$$\beta_t = \begin{bmatrix} \alpha \\ \gamma \end{bmatrix}$$

Kalman Filter: An Example

The initial condition (or posterior) at $t-1$ is

$$\mu_{t-1}|D_{t-1} \sim \mathcal{N}\left(b_{t-1}, Q_{t-1}\right)$$

where

$$D_t = \left\{r_t, D_{t-1}\right\}$$

is the information set available at t.

Kalman Filter: Prior

The prior at time t is

$$\mu_t | D_{t-1} \sim \mathcal{N}\left(a_t, P_t\right)$$

where

$$\begin{aligned}
E\left(\mu_t | D_{t-1}\right) &= E\left(c + \Phi\mu_{t-1} + u_t | D_{t-1}\right) \\
&= c + \Phi E\left(\mu_{t-1} | D_{t-1}\right) \\
&= \underbrace{c + \Phi b_{t-1}}_{a_t}
\end{aligned}$$

and

$$\begin{aligned}
Var\left(\mu_t | D_{t-1}\right) &= Var\left(c + \Phi\mu_{t-1} + u_t | D_{t-1}\right) \\
&= \Phi Var\left(\mu_{t-1} | D_{t-1}\right)\Phi' + \Sigma_{uu} \\
&= \underbrace{\Phi Q_{t-1}\Phi' + \Sigma_{uu}}_{P_t}
\end{aligned}$$

Kalman Filter: Prediction

The prediction at time t is

$$r_t | D_{t-1} \sim \mathcal{N}\left(f_t, S_t\right)$$

where

$$\begin{aligned}
E\left(r_t | D_{t-1}\right) &= E\left(Z_{t-1}\mu_{t-1} + X_{t-1}\beta + \varepsilon_t | D_{t-1}\right) \\
&= Z_{t-1} E\left(\mu_{t-1} | D_{t-1}\right) + X_{t-1}\beta \\
&= \underbrace{Z_{t-1} b_{t-1} + X_{t-1}\beta}_{f_t}
\end{aligned}$$

and

$$\begin{aligned}
Var\left(r_t | D_{t-1}\right) &= Var\left(Z_{t-1}\mu_{t-1} + X_{t-1}\beta + \varepsilon_t | D_{t-1}\right) \\
&= Z_{t-1} Var\left(\mu_{t-1} | D_{t-1}\right) Z'_{t-1} + \Sigma_{\varepsilon\varepsilon} \\
&= \underbrace{Z_{t-1} Q_{t-1} Z'_{t-1} + \Sigma_{\varepsilon\varepsilon}}_{S_t}
\end{aligned}$$

Kalman Filter: Joint Distribution

The joint distribution at time t is

$$\left(\begin{array}{c} r_t \\ \mu_t \end{array} \middle| D_{t-1} \right) \sim \mathcal{N}\left(\left[\begin{array}{c} f_t \\ a_t \end{array} \right], \left[\begin{array}{cc} S_t & G_t \\ G_t' & P_t \end{array} \right] \right).$$

where

$$\begin{aligned}
Cov\left(r_t, \mu_t | D_{t-1}\right) &= Cov\left(\begin{array}{c} Z_{t-1}\mu_{t-1} + X_{t-1}\beta + \varepsilon_t, \\ c + \Phi\mu_{t-1} + u_t | D_{t-1} \end{array} \right) \\
&= Z_{t-1} Cov\left(\mu_{t-1}, \mu_{t-1} | D_{t-1}\right) \Phi' \\
&= \underbrace{Z_{t-1} Q_{t-1} \Phi'}_{G_t}
\end{aligned}$$

Kalman Filter: Posterior

The posterior at time t is

$$\mu_t | D_t \sim \mathcal{N}\left(b_t, Q_t\right)$$

where

$$\begin{aligned} E\left(\mu_t | r_t, D_{t-1}\right) &= m_2 + V_{21} V_{11}^{-1}\left(X_1 - m_1\right) \\ &= \underbrace{a_t + G_t' S_t^{-1}\left(r_t - f_t\right)}_{b_t} \end{aligned}$$

and

$$\begin{aligned} Var\left(\mu_t | r_t, D_{t-1}\right) &= V_{22} - V_{21} V_{22}^{-1} V_{12} \\ &= \underbrace{P_t - G_t' S_t^{-1} G_t}_{Q_t} \end{aligned}$$

We use the conditional distribution results using $D_t = \{r_t, D_{t-1}\}$ so that

- $E\left(\mu_t | D_t\right) = E\left(\mu_t | r_t, D_{t-1}\right)$ and $Var\left(\mu_t | D_t\right) = Var\left(\mu_t | r_t, D_{t-1}\right)$.

Kalman Filter: Summary

Given

$$\mu_{t-1}|D_{t-1} \sim \mathcal{N}\left(b_{t-1}, Q_{t-1}\right) \quad \text{with } D_t = \{r_t, D_{t-1}\}$$

The **prediction equations** are

$$\begin{aligned}
E\left(\mu_t|D_{t-1}\right) &: \quad a_t = c + \Phi b_{t-1} \\
V\left(\mu_t|D_{t-1}\right) &: \quad P_t = \Phi Q'_{t-1}\Phi + \Sigma_{uu} \\
E\left(r_t|D_{t-1}\right) &: \quad f_t = Z_{t-1}b_{t-1} + X_{t-1}\beta \\
Var\left(r_t|D_{t-1}\right) &: \quad S_t = Z_{t-1}Q_{t-1}Z'_{t-1} + \Sigma_{\varepsilon\varepsilon} \\
Cov\left(r_t, \mu_t|D_{t-1}\right) &: \quad G_t = Z_{t-1}Q_{t-1}\Phi'.
\end{aligned}$$

The **updating equations** are

$$\begin{aligned}
E\left(\mu_t|r_t, D_{t-1}\right) &: \quad b_t = a_t + G'_t S_t^{-1}\left(r_t - f_t\right) \\
V\left(\mu_t|r_t, D_{t-1}\right) &: \quad Q_t = P_t - G'_t S_t^{-1} G_t.
\end{aligned}$$

Kalman Filter: Summary

Since the system is Gaussian, we have that

$$r_t | D_{t-1} \sim \mathcal{N}\left(f_t, S_t\right)$$

The sample log likelihood function is represented by

$$\ell\left(r|\theta\right) = -\frac{TN}{2}\ln\left(2\pi\right) - \frac{1}{2}\sum_{t=1}^{T}\ln\left|S_t\right| - \frac{1}{2}\sum_{t=1}^{T}\left(r_t - f_t\right)' S_t^{-1}\left(r_t - f_t\right)$$

where

- T is the number of observations,
- N is the number of parameters,
- θ refers to all unknown parameters.