# Empirical Finance: Methods & Applications

## Observed & Unobserved Predictors

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

$$
\begin{aligned}
P_t &= \frac{D_{t+1}}{R_{t+1}} + \frac{P_{t+1}}{R_{t+1}} \\
&= \frac{D_{t+1}}{R_{t+1}} + \frac{\frac{D_{t+2}}{R_{t+2}} + \frac{P_{t+2}}{R_{t+2}}}{R_{t+1}} \\
&= \frac{D_{t+1}}{R_{t+1}} + \frac{D_{t+2}}{R_{t+1}R_{t+2}} + \frac{P_{t+2}}{R_{t+1}R_{t+2}}
\end{aligned}
$$

# 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

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

# 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 $\rho$ 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?

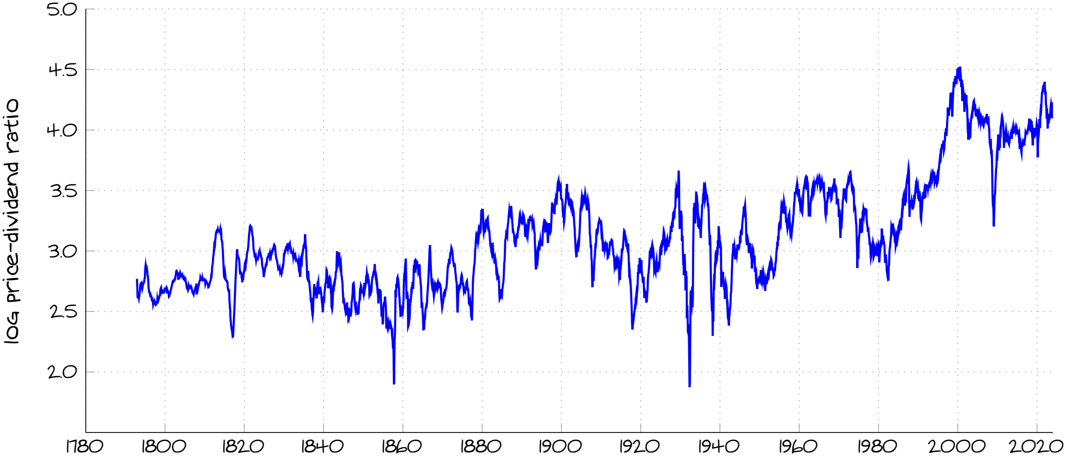
Decompose the variance of the log price-dividend ratio

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

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

- $pd_t$ must positively predict future $\Delta d_{t+j}$
- $pd_t$ must negatively predict future $r_{t+j}$

# 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 |
| $r_{t+1} = a_r + b_r(d_t - p_t) + \varepsilon_{t+1}^r$ | 0.097 | 1.92 | 4.0 | 4.0 |
| $\Delta d_{t+1} = a_d + b_d(d_t - p_t) + \varepsilon_{t+1}^{dp}$ | 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. $R_{t+1}^f$ 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?

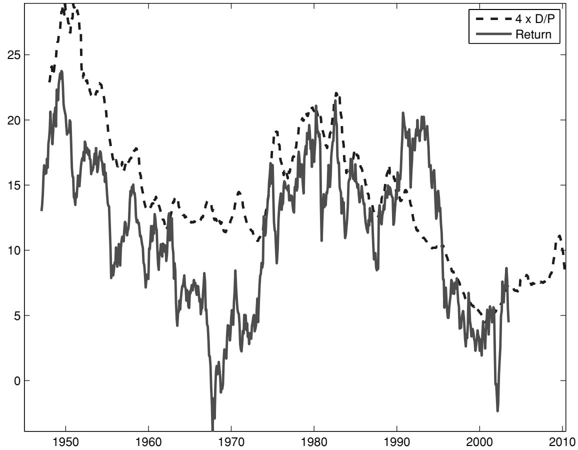



**Figure 1. Dividend yield and following 7-year return**. The dividend yield is multiplied by four. Both series use the CRSP value-weighted market index.

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 $\rho$ 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 $\Delta c_{t+i}$ 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?

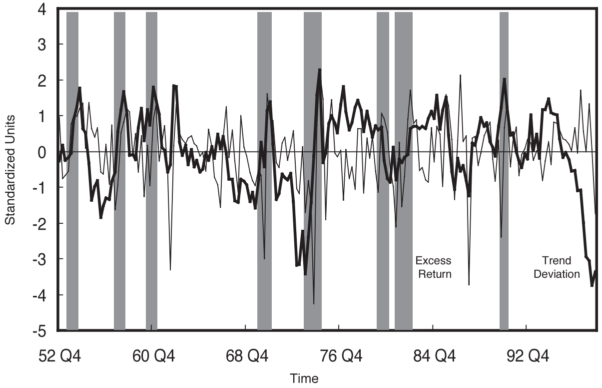



**Figure 1. Excess returns and trend deviations. Excess return is the return on the S&P Composite Index less the return on the three-month Treasury bill rate.** Trend deviation is the estimated deviation from the shared trend in consumption $c$, labor income $y$, and asset wealth $a$: $\widehat{cay}_t = c_t - \hat{\beta}_a a_t - \hat{\beta}_y y_t$. 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

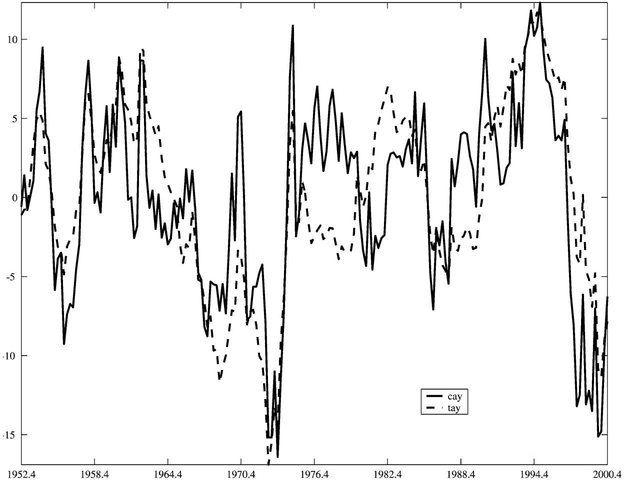
- The predictive power of $cay_t$ arises from a "look-ahead bias",
- This happens as the parameters $\beta_a$ and $\beta_y$ are fitted in-sample.

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,

- In-sample predictability does not translate into out-of-sample predictability,
- Almost all predictors fail to outperform the historical mean out of sample,
- They would have not have helped an investor with real-time information to earn a profit.

They show that the historical mean is a hard-to-beat benchmark model

# Bayesian Methods

# Bayesian Methods

Consider the probability rule for two random variables $A$ and $B$

joint probability of A and B

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

conditional probability of A given B

marginal probability of B

What is the probability that A takes a certain value and B a certain value at the same time?

If B is observed how likely is A?

What is the probability that B occurs, regardless of what happens to A?

# A Simple Example

Consider the following events

$$A \longrightarrow \text{It rains today}$$

$$B \longrightarrow \text{The sky is cloudy}$$

You have the following probabilities

$$p(A \mid B) = 0.75 \longrightarrow \text{If the sky is cloudy, it rains 75\% of the time}$$

$$p(B) = 0.40 \longrightarrow \text{40\% of days are cloudy}$$

Using the probability rule

$$p(A, B) = p(A \mid B)\, p(B) = 0.75 \times 0.40 = 0.30$$

A 30% probability that is both rainy and cloudy on the same day

# Bayesian Methods

Take the probability rule for two random variables $A$ and $B$

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

or its reverse

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

By combining them, we obtain the Bayes rule 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,
- $\theta$ refers to the parameters of a model,

The Bayes Theorem (or simply rule) states that

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

- $\theta$ is a random variable for *Bayesian econometrics* and not a fixed point,
- $p(\theta|y) \longrightarrow$ "given data $y$, what do we learn about the parameter $\theta$"?

Bayesian inference uses y to update our knowledge about $\theta$

# Bayesian Methods

Since we only care about $\theta$, we can simplify the Bayes rule as

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

posterior density of $\theta$

what we know about $\theta$
after observing the data

likelihood function

the density of the data
given the parameters

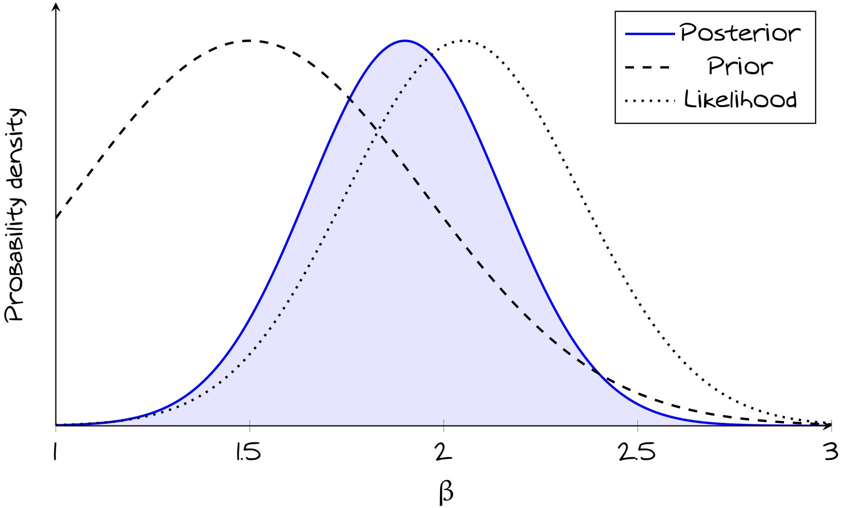
prior density of $\theta$

what we know about $\theta$
before observing the data

$p\,(y) \longrightarrow$ marginal density of $y$ (density of $y$ while ignoring information about $\theta$)

- A normalizing constant that can be ignored and $\propto$ means “proportional to“

# Bayesian Methods



# Why Ignoring p(y)?

$p\,(y) \longrightarrow$ marginal density of $y$ (likelihood of $y$ while ignoring $\theta$)

- A normalizing constant that can be ignored (not involving $\theta$),
- We then write $\propto$ (proportional to) instead of $=$ (equal) in the Bayes rule.

Think of the Bayes rule like grading a test

- The numerator is your score and the denominator is the maximum score,
- If everyone takes the same test (data $y$), the maximum score is the same for everyone,

$$\text{Percentage} = \frac{\text{Raw Score}}{\text{Maximum Score}}$$

To know who did best, comparing scores is enough.

Percentages are unnecessary, they simply rescale them.

# An Example: Who Is the Suspect?

## The data $y$

- A robbery was committed by someone wearing a Juventus Jersey.

## The parameter $\theta$

- There two suspects: A (known thief) and B (local baker).

## The prior $p(\theta)$

- Very high for suspect A $\longrightarrow p(\theta) = 10$,
- Quite low for suspect B $\longrightarrow p(\theta) = 1$.

## The marginal $p(y)$

- How common Juventus Jersey are is the same for both suspects.

# An Example: Who Is the Suspect?

## The likelihood $p(y \mid \theta)$

- Suspect A owns a Juventus Jersey $\longrightarrow p(y \mid \theta) = 1$,
- Suspect B owns a Juventus Jersey $\longrightarrow p(y \mid \theta) = 1$.

## The posterior $p(\theta|y)$

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

- Suspect A $\longrightarrow 10 \times 1 = 10$,
- Suspect B $\longrightarrow 1 \times 1 = 1$.

A is 10 times more likely to be the robber than B.

Dividing by the same number does not affect the comparison.

# Bayesian Methods

The Bayes rule states that

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

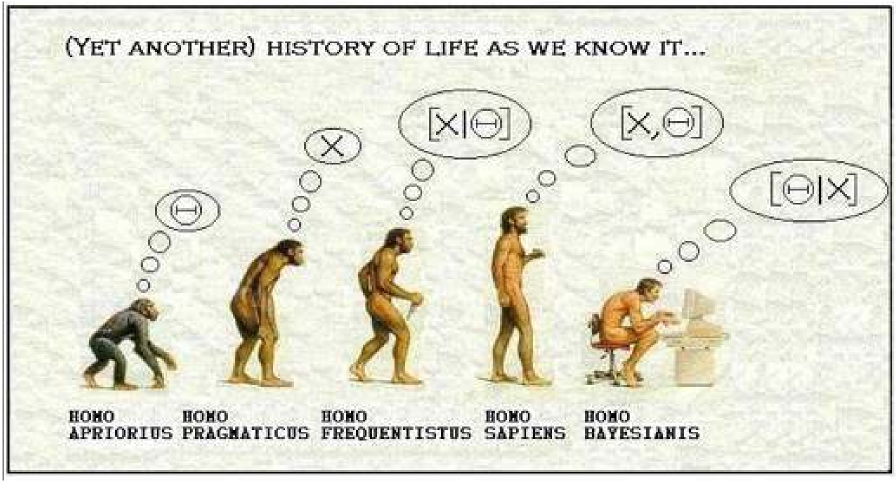
"The posterior is proportional to the likelihood times the prior".

Think of the Bayes rule as an updating rule

- We use data to update our prior views about $\theta$,
- The posterior combines prior beliefs with *current information*.

Prior Beliefs + Data = Posterior Beliefs

# Bayesian Methods



# Unobserved Factors

# Unobserved Factors

Many economic and financial variables are driven by forces that we cannot observe directly

- Asset prices may react to hidden state variables like liquidity, volatility, and risk aversion,
- A state space is used to model latent variables.

A state space model consists of

1. Measurement equation $\longrightarrow$ links actual data to latent or unobserved variables
2. State equation $\longrightarrow$ describes the evolution of the state variable.

The Kalman Filter (Kalman, 1960),

- An algorithm used to solve linear and Gaussian state space models.
- 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

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

# State Space: Examples

A predictive regression with observed $cay_{t-1}$ and unobserved $\mu_{t-1}$ predictors

Measurement Equation $\longleftarrow$ $r_t = \alpha + \gamma cay_{t-1} + \mu_{t-1} + \varepsilon_t$

State Equation $\longleftarrow$ $\mu_t = \phi\pi_{t-1} + \eta_t$

A predictive regression with time-varying slope coefficients $\gamma$ and $\delta$

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

State Equations

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

$$\delta_t = c_2 + \phi_2 \delta_{t-1} + u_t$$

You can also have contemporaneous regressions

# A State Space: A General Version

The measurement and state equations are defined as

M x 1 vector of time-invariant slopes

1 x M vector of regressors

1 x K vector of regressors

1 x 1 return

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

K x 1 vector of unobserved

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

K x 1 vector of constant

K x K matrix of parameters (with restricted eigenvalues)

# A State Space: A General Version

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

# Regression with an Unobserved Predictor

A predictive regression with observed $cay_{t-1}$ and unobserved $\mu_{t-1}$ predictors

$$\text{Measurement Equation} \longleftarrow \boxed{r_t} = \alpha + \gamma cay_{t-1} + \mu_{t-1} + \varepsilon_t$$

$$\text{State Equation} \longleftarrow \boxed{\mu_t} = \phi\pi_{t-1} + \eta_t$$

Nested into the general state space by setting

$$X_{t-1} = \begin{bmatrix} 1 & cay_{t-1} \end{bmatrix}$$

$$Z_{t-1} = 1$$

$$\Phi = \phi$$

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

# Regression with Time-Varying Parameters

A predictive regression with time-varying slope coefficients $\gamma$ and $\delta$

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

State Equations
$$\gamma_t = c_1 + \phi_1 \gamma_{t-1} + \eta_t$$
$$\delta_t = c_2 + \phi_2 \delta_{t-1} + u_t$$

Nested into the general state space by setting

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

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

# Kalman Filter

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

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

is the information set available at $t$.

# Kalman Filter: Prior

The prior at time $t$ is

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

where

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

and

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

# 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

$$\begin{pmatrix} r_t \\ \mu_t \end{pmatrix} | D_{t-1} \sim \mathcal{N}\left( \begin{bmatrix} f_t \\ a_t \end{bmatrix}, \begin{bmatrix} S_t & G_t \\ G_t' & P_t \end{bmatrix} \right).$$

where

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

# 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

$$\begin{aligned} Var\left(\mu_t | r_t, D_{t-1}\right) &= V_{22} - V_{21} V_{11}^{-1} V_{12} \\ &= \underbrace{P_t - G_t' S_t^{-1} G_t}_{Q_t} \end{aligned}$$

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

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

The sample log likelihood (sum of per-period log-density) is

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

where

- $T$ is the number of observations,
- $N$ is the number of parameters,
- $\theta$ refers to all unknown parameters.

# Likelihood Function

```
function kalman_lik(theta)

# Extract parameters to be estimated
alpha = theta[1]
beta  = theta[2]
var_e = theta[3]
phi   = theta[4]
var_u = theta[5]
T     = size(r, 1)

# Initialize estimates
b_0 = 0
Q_0 = var_u/(1-phi^2)

#Initial values for Kalman filter
b[1] = b_0
Q[1] = Q_0
```

# Likelihood Function

```
# Store estimates
L   = zeros(T)         # Store log-likelihood sum
a   = zeros(T)
f   = zeros(T)
b   = zeros(T)
P   = zeros(T)
S   = zeros(T)
G   = zeros(T)
Q   = zeros(T)
```

# Likelihood Function

```
    # Kalman filtering loop
    for t in 2:T

    # Prediction step
    a[t] = phi*b[t-1]
    f[t] = b[t-1] + (alpha + x[t-1]*beta)
    P[t] = (phi^2)*Q[t-1] + var_u
    S[t] = Q[t-1] + var_e

    if S[t] <= 0
    S[t] = max(S[t], 1e-8) # Avoid log of zero or negative variance
    end

    G[t] = Q[t-1]*phi

    # Updating step
    b[t] = a[t] + (G[t] / S[t] ) * (r[t] - f[t])
    Q[t] = P[t] - (G[t]^2 / S[t])

```

# Likelihood Function

```
    # Compute log-likelihood (Gaussian)
    L[t] = -0.5*log(2*pi) - 0.5*log(S[t]) - 0.5*((r[t] - f[t])^2 / S[t])
    end

    # Store beta_hat globally to access it after optimization
    global f_final = f
    global S_final = S

    return -sum(L)  # Return negative log-likelihood for minimization

    end


```

# Likelihood Function

```
    # Compute log-likelihood (Gaussian)
    L[t] = -0.5*log(2*pi) - 0.5*log(S[t]) - 0.5*((r[t] - f[t])^2 / S[t])
    end

    # Store beta_hat globally to access it after optimization
    global f_final = f
    global S_final = S

    return -sum(L)  # Return negative log-likelihood for minimization

    end
```

# Optimizing the Likelihood Function

```
# Load Data
data = CSV.read("output/simulated_predictor.csv", DataFrame)

# setup the data
r = data.y   #returns
x = data.x   # observable predictor
u = data.u   # unobservable predictor
T = size(data, 1)

# ==============================
# Maximum Likelihood Estimation (MLE)
# ==============================
initial_prm = [0, 0.0, 0.01, 0.90, 0.01]  # Initial guesses for [beta, var_e, c, phi, var_u]
opt_result  = optimize(kalman_lik, initial_prm, NelderMead())

# Extract estimated parameters
alpha_mle, beta_mle, var_e_mle, phi_mle, var_u_mle = Optim.minimizer(opt_result)
```

# Finale Estimates