Empirical Finance: Week 4 Comprehensive Exam Guide

Observed & Unobserved Predictors

With Professor's Teaching Points & Exam Instructions

Extracted from Lecture Captions & Handout 4

Exam Material: Weeks 1–6 Only

CRITICAL EXAM INSTRUCTIONS

Exam Scope: Material up to Week 6 only. Week 7 is optional and NOT examined.

UK Grading: 70% = Distinction; Target average = 66–68%

Professor's Rule: "Unless I tell you explicitly in class that you have to remember something for the exam, you're not supposed to."

Kalman Filter Question: "Pay attention because one of the questions, **probably worth 30 points**, I might give you a system and ask you to derive the prior or the posterior."

Minor Errors Policy: "Any minor errors? Don't worry. I'm not going to count for the final mark."

Contents

1 Campbell-Shiller Decomposition

EXAM GUIDANCE – Professor's Explicit Statements

- "In the exam I'm not going to ask you to **derive all the steps**"
- "There could be questions, especially in **multiple choice**, when I say what's the relationship between price dividend, future stock returns and dividend growth"
- "You have to remember the **SIGN of the correlation**"
- "You don't need to remember the full derivation"

1.1 Professor's Key Teaching Points

Professor's Intuition

Why do we need this?

- The price-dividend relationship is **nonlinear** – we cannot test it directly
- "We started from a simple identity, the identity we have seen in the first line of lecture one"
- "Then we said, let's make this term stationary by working with the price dividend ratio"
- "This is not even a model because there is no equilibrium condition. This is an **identity**. So we started from an identity and we have derived a predictor."
- "Maybe the most popular predictor in the equity space which is the price dividend ratio"
- "The price dividend doesn't come from the model. It's just an identity. So it should hold in the data."

1.2 Formulas to MEMORISE

MUST MEMORISE – Correlation Signs

KEY RELATIONSHIP – The log price-dividend ratio pd_t is:

- **POSITIVELY** correlated with future dividend growth Δd_{t+j}
- **NEGATIVELY** correlated with future stock returns r_{t+j}

Memory Aid:

$$pd_t = \frac{\kappa}{1-\rho} + \underbrace{\sum_{j=0}^{\infty} \rho^j \Delta d_{t+j+1}}_{(+)\text{ correlation}} - \underbrace{\sum_{j=0}^{\infty} \rho^j r_{t+j+1}}_{(-)\text{ correlation}}$$

WARNING – Common Exam Trap (70% Failed This Question Last Year)

Professor warns: "Don't learn things mechanically. I could make a small change in the exam and trick you."

The Trap: If asked about **DIVIDEND-PRICE** ($d_t - p_t$) instead of **PRICE-DIVIDEND** ($p_t - d_t$), the signs **REVERSE**:

- $d_t - p_t$ is **NEGATIVELY** correlated with future dividend growth
- $d_t - p_t$ is **POSITIVELY** correlated with future stock returns

"Some papers, what they do, rather than taking p minus d, they take d minus p. If you work with $d - p$ then you have a negative relationship with dividend growth, positive relationship with future stock returns."

Try to understand what the formula suggests, not just memorise!

1.3 Formulas Given on Exam (Know How to Apply)

Campbell-Shiller Log-Linear Approximation:

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

where:

$$\kappa = \ln(1 + e^{\overline{pd}}) - \frac{e^{\overline{pd}}}{1 + e^{\overline{pd}}} \cdot \overline{pd} \tag{2}$$

$$\rho = \frac{e^{\overline{pd}}}{1 + e^{\overline{pd}}} \approx 0.964 \text{ (annual US data)} \tag{3}$$

Present-Value Relationship:

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

Variance Decomposition:

$$\text{Var}(pd_t) = \text{Cov}\left(pd_t, \sum_{j=0}^{\infty} \rho^j \Delta d_{t+j+1}\right) - \text{Cov}\left(pd_t, \sum_{j=0}^{\infty} \rho^j r_{t+j+1}\right) \tag{5}$$

Professor's Intuition – Why This Matters

"If the log price-dividend ratio varies over time, then pd_t must either predict future dividend growth (positive covariance) OR predict future returns (negative covariance). The identity must hold – it's an accounting relationship, not a hypothesis."

"All variation in the log price dividend ratio must be explained either by this beta or by this beta. That's why we derive this identity in variance and covariance form."

1.4 Empirical Evidence (Cochrane 2008, 2011)

Regression	b	t-stat	R^2
$r_{t+1} = a + b(d_t - p_t) + \varepsilon_{t+1}$	0.097	1.92	4.0%
$\Delta d_{t+1} = a + b(d_t - p_t) + \varepsilon_{t+1}$	0.008	0.18	0.0%

Key Finding

Dividend-price ratio predicts RETURNS, NOT dividend growth.
The variation in pd_t is almost entirely due to **time-varying expected returns** (discount rates), not cash flow news. This is why Cochrane (2011) titled his presidential address "*Discount Rates*."

"In this paper, he argues that in theory and in the data, the dividend price is a good predictor should be a good predictor for future stock returns."

2 Consumption-Wealth Ratio: *cay*

2.1 Professor's Key Teaching Points

Professor's Intuition

- Similar logic to Campbell-Shiller but applied to consumption-wealth relationship
- Aggregate wealth is **unobservable** – need to proxy with asset holdings and labor income
- Lettau & Ludvigson (2001) is the seminal paper
- The *cay* variable should predict returns if consumption-wealth ratio varies

"If wealth is high relative to consumption, expected returns may be low (or vice versa), so *cay* can forecast returns."

2.2 Formulas Given on Exam (Know How to Apply)

Log Consumption-Wealth Ratio:

$$c_t - w_t = E_t\left[\sum_{i=1}^{\infty} \rho^i (r_{t+i} - \Delta c_{t+i})\right] \tag{6}$$

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

Practical Implementation (since W_t is unobservable):

$$\widehat{cay}_t = c_t - \omega \cdot a_t - (1-\omega) \cdot y_t \tag{7}$$

where:

- c_t = log consumption
- a_t = log asset holdings
- y_t = log labor income
- ω = share of asset wealth in total wealth

Estimation via Dynamic OLS (DOLS):

$$c_t = \alpha + \beta_a a_t + \beta_y y_t + \sum_{i=-k}^{k} b_{a,i}\Delta a_{t-i} + \sum_{i=-k}^{k} b_{y,i}\Delta y_{t-i} + \varepsilon_t \tag{8}$$

Then: $\widehat{cay}_t = c_t - \hat{\beta}_a a_t - \hat{\beta}_y y_t$

2.3 Empirical Evidence

In-Sample Results (Lettau & Ludvigson 2001):

- *cay* predicts stock returns: $R^2 \approx 9$–10%, t-stat ≈ 2.6–3.2
- *cay* does NOT predict dividend growth: $R^2 \approx 0\%$, t-stat ≈ 0.18

Out-of-Sample Controversy

Brennan & Xia (2005) "*tay* vs *cay*" paper argues:

- *cay*'s predictive power arises from **look-ahead bias**
- Replacing consumption with a linear trend (*tay*) works equally well in-sample
- **Both *cay* and *tay* lose their out-of-sample forecasting power** when re-estimated every period

Welch & Goyal (2008): The historical mean is a hard-to-beat benchmark out of sample.

3 Bayesian Methods

3.1 Professor's Key Teaching Points

Professor's Intuition

- Bayesian inference treats θ as a **random variable**, not a fixed point
- We use data to **update our prior beliefs** about parameters
- The posterior combines prior information with current data
- "Most of the time you are Bayesian, even though you think that you're not Bayesian"
- "In standard econometrics, you only work with the likelihood. You ignore the prior. The truth is that before we deal with data, we always have priors."

3.2 Formulas to MEMORISE

MUST MEMORISE – Bayes Rule

Bayes Rule:

$$p(\theta|y) = \frac{p(y|\theta) \cdot p(\theta)}{p(y)} \quad \Longrightarrow \quad \boxed{p(\theta|y) \propto p(y|\theta) \cdot p(\theta)} \tag{9}$$

In words:

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}$$

Or equivalently:

$$\text{Prior Beliefs} + \text{Data} = \text{Posterior Beliefs}$$

3.3 Understanding the Components

Term	Meaning
$p(\theta\|y)$	**Posterior:** What we know about θ AFTER observing the data
$p(y\|\theta)$	**Likelihood:** Density of the data given the parameters
$p(\theta)$	**Prior:** What we know about θ BEFORE observing the data
$p(y)$	**Marginal:** Normalising constant (can be ignored for comparison)

Professor's Intuition – Why Can We Ignore $p(y)$?

Test Grading Analogy:

- The numerator is your **raw score**
- The denominator is the **maximum score**
- If everyone takes the same test (same data y), the maximum is the same for everyone
- To compare who did best, comparing raw scores is enough
- Percentages (dividing by $p(y)$) just rescale – they don't change the ranking

"To rank you I don't need the maximum score. It's basically this is $p(y)$."

Professor's Example – Robbery Suspect

Data y: A robbery was committed by someone wearing a Juventus jersey.

Two suspects:

- Suspect A (known thief): Prior $p(\theta_A) = 10$ (very likely)
- Suspect B (local baker): Prior $p(\theta_B) = 1$ (less likely)

Likelihood: Both own Juventus jerseys $\Rightarrow p(y|\theta) = 1$ for both.

Posterior comparison:

- Suspect A: $10 \times 1 = 10$
- Suspect B: $1 \times 1 = 1$

$\Rightarrow$ **A is 10 times more likely to be the robber. No need to compute $p(y)$!**

"I don't need $p(y)$ because it's common to both of them. So I consider that the posterior is basically the likelihood times the prior."

4 Kalman Filter

EXAM GUIDANCE – Professor's Explicit Statements (HIGH PROBABILITY QUESTION)

- "In the exam, I might ask you the Kalman filter"
- "I might give you a **simple specification** and ask you calculate the prior, fill up the distribution"
- "You just need to remember how to calculate the **expected value and the variance**"
- "You **DON'T need to memorise** because I can give you a completely different specification"
- "I want you to remember the **STEPS**"
- "If you remember the steps, you're fine"
- "Any minor errors? Don't worry. I'm not going to count for the final mark"

EXAM QUESTION HINT: "Pay attention because one of the questions, **probably worth 30 points**, I might give you a system and ask you to derive the prior or the posterior."

4.1 Why Use the Kalman Filter?

Professor's Intuition – The Core Problem

"Returns are basically **white noise**. Noise dominates the signal. Signal-to-noise ratio is very low. If I regress r_t on observables, all betas are insignificant."

Solution: Extract the predictable component r_t^* (unobserved), then regress r_t^* on macro variables.

"The Kalman filter extracts the juice from the data."

"You are running a fishing expedition. You have your set of asset returns. You have no idea on how to explain those returns. You want to build up a story. So you extract the juice, and then you taste the juice, and then you decide whether you want to make a bigger bottle out of your experiment."

4.2 State Space Model Structure

Measurement Equation:

$$r_t = X_{t-1}\beta + Z_{t-1}\mu_{t-1} + \varepsilon_t \tag{10}$$

State Equation:

$$\mu_t = c + \Phi\mu_{t-1} + u_t \tag{11}$$

Error Structure:

$$\begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{\varepsilon\varepsilon} & 0 \\ 0 & \Sigma_{uu} \end{pmatrix}\right) \tag{12}$$

4.3 Kalman Filter STEPS to Remember

MUST REMEMBER – The Three Steps

Given: $\mu_{t-1}|D_{t-1} \sim N(b_{t-1}, Q_{t-1})$ where $D_t = \{r_t, D_{t-1}\}$

STEP 1: PREDICTION (Prior at time t)

$$E(\mu_t|D_{t-1}) = a_t = c + \Phi b_{t-1} \tag{13}$$
$$\text{Var}(\mu_t|D_{t-1}) = P_t = \Phi Q_{t-1}\Phi' + \Sigma_{uu} \tag{14}$$

STEP 2: FORECAST

$$E(r_t|D_{t-1}) = f_t = Z_{t-1}b_{t-1} + X_{t-1}\beta \tag{15}$$
$$\text{Var}(r_t|D_{t-1}) = S_t = Z_{t-1}Q_{t-1}Z'_{t-1} + \Sigma_{\varepsilon\varepsilon} \tag{16}$$
$$\text{Cov}(r_t, \mu_t|D_{t-1}) = G_t = Z_{t-1}Q_{t-1}\Phi' \tag{17}$$

STEP 3: UPDATING (Posterior at time t)

$$E(\mu_t|D_t) = b_t = a_t + G'_tS_t^{-1}(r_t - f_t) \tag{18}$$
$$\text{Var}(\mu_t|D_t) = Q_t = P_t - G'_tS_t^{-1}G_t \tag{19}$$

Professor's Summary of the Logic

"We have the same logic. We have prior. We have the likelihood and we have the posterior. Then the posterior becomes the prior if we have the data and then prior, likelihood, posterior. That's the way it works."

The Loop:

$$\text{Prior} \rightarrow \text{Prediction} \rightarrow \text{Observe Data} \rightarrow \text{Update} \rightarrow \text{New Prior} \rightarrow \dots$$

"So when you estimate the parameters of a linear regression maximum likelihood, you have that the mean and the variance are fixed through time. In this case they change over time."

4.4 Log-Likelihood Function

Since the system is Gaussian: $r_t|D_{t-1} \sim N(f_t, S_t)$

$$\ell(r|\theta) = \sum_{t=1}^{T}\left[-\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|S_t| - \frac{1}{2}(r_t - f_t)'S_t^{-1}(r_t - f_t)\right] \tag{20}$$

4.5 Special Cases (Examples from Slides)

1. Regression with Unobserved Predictor:

$$r_t = \alpha + \gamma \cdot cay_{t-1} + \mu_{t-1} + \varepsilon_t \tag{21}$$
$$\mu_t = \phi\mu_{t-1} + \eta_t \tag{22}$$

2. Regression with Time-Varying Parameters:

$$r_t = \alpha + \gamma_{t-1} \cdot dp_{t-1} + \delta_{t-1} \cdot cay_{t-1} + \varepsilon_t \tag{23}$$
$$\gamma_t = c_1 + \phi_1 \gamma_{t-1} + \eta_t \tag{24}$$
$$\delta_t = c_2 + \phi_2 \delta_{t-1} + u_t \tag{25}$$

Professor's Practical Advice

"I've shown you how to derive the Kalman filter. So if I give you a different specification, in principle you should be able to derive it."

"If you expect to stand below the tree waiting for the apple to fall in your mouth, forget it. The real world doesn't work like that. **You have to climb the tree and grab the apple.**"

5 Quick Reference: Exam Checklist

Topic	MEMORISE	UNDERSTAND (Given)
Campbell-Shiller	Signs of correlations: pd_t vs returns (**negative**) pd_t vs div growth (**positive**)	Full derivation, κ and ρ formulas
cay	Basic formula: $cay_t = c_t - \omega a_t - (1-\omega)y_t$	DOLS estimation, out-of-sample issues
Bayes Rule	$p(\theta\|y) \propto p(y\|\theta) \cdot p(\theta)$	Why ignore $p(y)$
Kalman Filter	**The STEPS:** Prediction $\rightarrow$ Forecast $\rightarrow$ Update	Specific formulas will be given; apply to new specification

Formulas You MUST Memorise

1. Campbell-Shiller Correlations:

- pd_t is **POSITIVELY** correlated with future dividend growth
- pd_t is **NEGATIVELY** correlated with future stock returns
- **WATCH OUT:** If $d_t - p_t$ (dividend-price), signs REVERSE!

2. Bayes Rule:

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}$$

3. Kalman Filter Steps (in words):

1. **PREDICTION:** Use state equation to predict μ_t given past info
2. **FORECAST:** Use measurement equation to forecast r_t
3. **UPDATE:** Combine prediction with new observation to get posterior

Conceptual Understanding Required

- **Campbell-Shiller:** Why this is an identity not a model; why we need log-linearisation
- ***cay*:** Brennan & Xia critique (look-ahead bias); in-sample vs out-of-sample
- **Bayes:** Why we can ignore $p(y)$ (normalising constant)
- **Kalman Filter:** Why we need it (extract signal from noise); how to apply steps to new specification

Professor's Final Warnings

- "Don't learn things mechanically. I could make a small change in the exam and trick you."
- "Try to keep up with material on a weekly basis."
- "Try to understand what the formula suggests."
- "Make my life easy when marking – be clear, don't mess up."
- "Minor differences in numbers won't be penalised."
- "If you stand below the tree waiting for the apple to fall, forget it. Climb the tree and grab the apple."