Problem Set 2: Solutions

State Space, Trends, and Policy Regime Change

Rafael Lincoln

Econometrics II – NYU PhD Program

Spring 2026

Today's Roadmap

Problem Set 2 Solutions

1. **Problem 1: Long-Run Risk (Bansal–Yaron)**
   - VAR representation, stationarity, Kalman filter
2. **Problem 2: Beveridge–Nelson Trend**
   - VAR-based trend extraction, random walk property
3. **Problem 3: Clarida–Galí–Gertler NK Model**
   - AR(1) derivation, LR/Wald/LM tests, Lucas critique
4. **Problem 4: Kalman Filter with Correlated Innovations**
5. **Problem 5: ARMA(2,2) to State Space**

Problem 1: Long-Run Risk

Bansal–Yaron (2004) Model

Problem 1: Setup

Model (Bansal–Yaron long-run risk):

$$x_t = \rho x_{t-1} + \sigma_\varepsilon \varepsilon_t$$
$$g_t = x_t + \sigma_\eta \eta_t$$

where $(\varepsilon_t, \eta_t)' \sim$ i.i.d. $\mathcal{N}(0, I_2)$.

Parameters: $\theta = (\rho, \sigma_\varepsilon, \sigma_\eta)'$

Interpretation:

- x_t: latent expected growth rate ("long-run risk")
- g_t: observed consumption/dividend growth
- The latent x_t creates persistent predictability in growth

Problem 1(a): VAR Representation

Question: *Derive a VAR representation for $(x_t, g_t)'$. What condition on θ is necessary for covariance stationarity?*
Solution: Stack the equations:

$$\underbrace{\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}}_{B} \begin{pmatrix} x_t \\ g_t \end{pmatrix} = \underbrace{\begin{pmatrix} \rho & 0 \\ 0 & 0 \end{pmatrix}}_{A} \begin{pmatrix} x_{t-1} \\ g_{t-1} \end{pmatrix} + \underbrace{\begin{pmatrix} \sigma_\varepsilon & 0 \\ 0 & \sigma_\eta \end{pmatrix}}_{C} \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix}$$

Premultiply by $B^{-1} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$:

$$\begin{pmatrix} x_t \\ g_t \end{pmatrix} = \begin{pmatrix} \rho & 0 \\ \rho & 0 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ g_{t-1} \end{pmatrix} + \begin{pmatrix} \sigma_\varepsilon & 0 \\ \sigma_\varepsilon & \sigma_\eta \end{pmatrix} \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix}$$

Stationarity Condition

Eigenvalues of $B^{-1}A$ are $\{\rho, 0\}$. **Stationarity requires** $|\rho| < 1$.

Problem 1(a): Stationarity — Two Perspectives

Why $|\rho| < 1$? Two equivalent ways to see this:

1. Characteristic polynomial approach:

- Stationarity requires all roots of the characteristic polynomial $\det(I - \Phi z) = 0$ to lie **outside** the unit circle (equivalently, eigenvalues of Φ inside unit circle)
- For our VAR: $\det \begin{pmatrix} 1 - \rho z & 0 \\ -\rho z & 1 \end{pmatrix} = (1 - \rho z) \cdot 1 = 0$
- Root: $z = 1/\rho$. For $|z| > 1$, need $|\rho| < 1$

2. Economic intuition:

- $x_t = \rho x_{t-1} + \sigma_\varepsilon \varepsilon_t$ is AR(1) — stationary iff $|\rho| < 1$
- $g_t = x_t + \sigma_\eta \eta_t$ is just x_t plus i.i.d. noise
- If x_t is stationary, then g_t must also be stationary
- The "extra noise" η_t doesn't affect persistence, only variance

Problem 1(b): Likelihood for Dependent Data — Review

Question: *Assuming stationarity and that x_t and g_t are both observable, explain how you would estimate θ by maximum likelihood.*

What we need:

- For **dependent data**, the joint density is not a product of i.i.d. marginals.
- We use the **decomposition of the log-likelihood** (with $\mathcal{F}_{t-1}^Y = \sigma(Y_0, \dots, Y_{t-1})$, process adapted to Y):

$$\log f(Y_0, Y_1, \dots, Y_T) \; = \; \log f(Y_0) + \sum_{t=1}^{T} \log f(Y_t \mid \mathcal{F}_{t-1}^Y)$$

VAR(1) case:

- $Y_t = \Phi Y_{t-1} + u_t, \quad u_t \sim \mathcal{N}(0, \Sigma_u)$.
- By **Markov property**: $Y_t \mid \mathcal{F}_{t-1}^Y$ depends only on Y_{t-1}, and is Gaussian:

$$Y_t \mid \mathcal{F}_{t-1}^Y \; \sim \; \mathcal{N}(\Phi Y_{t-1}, \, \Sigma_u)$$

- Write the log-likelihood (including $\log f(Y_0)$) and maximize it $\Rightarrow$ next slide.

Problem 1(b): Solution — MLE = OLS

From the decomposition: Let $Y_t = (x_t, g_t)'$. The reduced-form VAR is $Y_t = \Phi Y_{t-1} + u_t$, $u_t \sim \mathcal{N}(0, \Sigma_u)$.
Log-likelihood: $\ell_T(\theta) = \log f(Y_0) + \sum_{t=1}^{T} \log f(Y_t \mid \mathcal{F}_{t-1}^{Y})$. For the Gaussian VAR, the conditional terms give:

$$\ell_T(\theta) = \log f(Y_0) - \frac{T}{2} \log |\Sigma_u| - \frac{1}{2} \sum_{t=1}^{T} (Y_t - \Phi Y_{t-1})' \Sigma_u^{-1} (Y_t - \Phi Y_{t-1})$$

Maximizing w.r.t. Φ: the term that depends on Φ is the sum of squared (generalized) residuals. Under Gaussianity, the **MLE for Φ is OLS** equation-by-equation:

$$\hat{\Phi} = \left(\sum_{t=1}^{T} Y_t Y_{t-1}' \right) \left(\sum_{t=1}^{T} Y_{t-1} Y_{t-1}' \right)^{-1}$$

Then back out $(\rho, \sigma_\varepsilon, \sigma_\eta)$ from $\hat{\Phi}$ and $\hat{\Sigma}_u$.

Problem 1(c): Review — From state-space model to likelihood

Question: *Now assume that g_t is observable but x_t is not. Assuming stationarity, explain how you would estimate θ by maximum likelihood.*

Review:

- Besides this problem, the same ideas matter in a **broader context**: **estimation of structural models**.
- Classical ML and **Bayesian** estimation both rely on likelihood evaluation.
- Why useful? We use it when estimating **structural models** (e.g. business-cycle models with unobserved states).

From state-space model to likelihood: A model is a parameter vector ψ giving rise to a state-space system for observables y_t:

- **Observation equation**: $y_t = \Psi(s_t; \psi) + u_t, \quad u_t \sim F_u(\cdot; \psi)$
- **State transition**: $s_t = \Phi(s_{t-1}, \varepsilon_t; \psi), \quad \varepsilon_t \sim F_\varepsilon(\cdot; \psi)$

Q: What is the likelihood of $y_{1:T}$ given ψ? **Standard approach: filtering.**

Problem 1(c): Review — Likelihood factorization and evaluation

Likelihood factorization ($\mathcal{F}_{t-1}^{Y} = \sigma(y_0, \ldots, y_{t-1})$):

$$p(y_{1:T} \mid \psi) \;=\; \prod_{t=1}^{T} p(y_t \mid \mathcal{F}_{t-1}^{Y}, \psi)$$

0. Initial: $p(s_0 \mid y_{1:0})$ (e.g. stationary).

1. Forecasting (predict s_t, then y_t, given past):

- **(a) Transition equation:**

$$p(s_t \mid y_{1:t-1}) \;=\; \int p(s_t \mid s_{t-1}, y_{1:t-1})\, p(s_{t-1} \mid y_{1:t-1})\, ds_{t-1}$$

- **(b) Measurement equation:**

$$p(y_t \mid y_{1:t-1}) \;=\; \int p(y_t \mid s_t, y_{1:t-1})\, p(s_t \mid y_{1:t-1})\, ds_t$$

2. Updating (Bayes):

$$p(s_t \mid y_{1:t}) \;=\; \frac{p(y_t \mid s_t, y_{1:t-1})\, p(s_t \mid y_{1:t-1})}{p(y_t \mid y_{1:t-1})}$$

Problem 1(c): Review — Kalman filter (Gaussian case)

When the model is linear and shocks are Gaussian, we get the **Kalman filter**—and everything becomes easy.

Likelihood: Chain rule $p(y^T) = p(y_T \mid y^{T-1}) \cdots p(y_2 \mid y_1)\, p(y_1)$, with $y^t = (y_t, \ldots, y_1)$. Under normality:

$$y_t \mid \mathcal{F}^Y_{t-1} \;\sim\; \mathcal{N}(\hat{y}_{t|t-1},\, V_{t|t-1})$$

Definitions (linear model $y_t = b_t + Z_t s_t + u_t$, $s_t = A_t s_{t-1} + B_t \varepsilon_t$):

$$\begin{aligned}
\hat{y}_{t|s} &\equiv \mathbb{E}[y_t \mid y^s] \;=\; b_t + Z_t \hat{s}_{t|s} \\
V_{t|s} &\equiv \mathrm{Var}(y_t \mid y^s) \;=\; Z_t P_{t|s} Z_t' + H_t \\
\hat{s}_{t|s} &\equiv \mathbb{E}[s_t \mid y^s], \qquad P_{t|s} \equiv \mathrm{Var}(s_t \mid y^s)
\end{aligned}$$

with $H_t = \mathrm{Var}(u_t)$, $Q_t = \mathrm{Var}(\varepsilon_t)$.

If we have $\hat{s}_{t|t-1}$ and $P_{t|t-1}$, we can compute the **log-likelihood**. These objects are computed recursively from $\hat{s}_{0|0}$, $P_{0|0}$.

Problem 1(c): Review — The Kalman filter

- If we can compute $\hat{s}_{t|t-1}$ and $P_{t|t-1}$, we can compute the log likelihood.
    - We compute these **recursively**, starting from $\hat{s}_{0|0}$ and $P_{0|0}$.

Forecasting step. Use $\hat{s}_{t-1|t-1}$, $P_{t-1|t-1}$ and the transition equation to create forecasts:

$$\hat{s}_{t|t-1} = A_t \hat{s}_{t-1|t-1}, \qquad P_{t|t-1} = A_t P_{t-1|t-1} A_t' + B_t Q_t B_t' \quad (Q_t = \mathrm{Var}(\varepsilon_t))$$

Observation forecast: $\hat{y}_{t|t-1} = b_t + Z_t \hat{s}_{t|t-1}$, $\quad V_{t|t-1} = Z_t P_{t|t-1} Z_t' + H_t$.
Updating step. Update $\hat{s}_{t|t-1}$ and $P_{t|t-1}$ to condition on y_t.

- Forecast error: $\tilde{y}_t = y_t - \hat{y}_{t|t-1}$.
- Kalman gain: $K_t = P_{t|t-1} Z_t' V_{t|t-1}^{-1}$.
- Update $\hat{s}_{t|t} = \hat{s}_{t|t-1} + K_t \tilde{y}_t$.
- Update $P_{t|t} = (I - K_t Z_t) P_{t|t-1}$.

Log-likelihood (scalar): $\ell_T(\psi) = -\frac{1}{2} \sum_{t=1}^{T} \left[\log V_{t|t-1} + \tilde{y}_t^2 / V_{t|t-1} \right] + \text{const}$.

Problem 1(c): MLE When x_t Latent — State Space

Question: *Now assume that g_t is observable but x_t is not. Assuming stationarity, explain how you would estimate θ by maximum likelihood.*

Solution. Our model is a **linear Gaussian state-space**:

- **State equation**: $x_t = \rho x_{t-1} + \sigma_\varepsilon \varepsilon_t$
- **Observation equation**: $g_t = x_t + \sigma_\eta \eta_t$

Mapping to review notation ($y_t = g_t$, $s_t = x_t$, $\psi = \theta$):

- Observation: $b_t = 0$, $Z_t = 1$, $H_t = \sigma_\eta^2$
- Transition: $A_t = \rho$, $B_t = \sigma_\varepsilon$, $Q_t = 1$

Estimation strategy:

1. Run the **Kalman filter** to get $\hat{x}_{t|t}$ and prediction errors.
2. Evaluate the **likelihood** via prediction error decomposition.
3. **Maximize** $\ell_T(\theta)$ numerically over $\theta = (\rho, \sigma_\varepsilon, \sigma_\eta)$.

Problem 1(c): Kalman Filter Recursions

Initialization: $\hat{x}_{0|0} = 0$, $P_{0|0} = \frac{\sigma_\varepsilon^2}{1-\rho^2}$ (unconditional variance).
For $t = 1, \ldots, T$:
Prediction step:

$$\hat{x}_{t|t-1} = \rho \hat{x}_{t-1|t-1}, \quad P_{t|t-1} = \rho^2 P_{t-1|t-1} + \sigma_\varepsilon^2$$

Update step (upon observing g_t):

- **Prediction error**: $v_t = g_t - \hat{x}_{t|t-1}$
- **Prediction error variance**: $F_t = P_{t|t-1} + \sigma_\eta^2$ (innovation variance)
- **Kalman gain**: $K_t = P_{t|t-1}/F_t$
- **Updated state estimate**: $\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t v_t$
- **Updated state variance**: $P_{t|t} = (1 - K_t) P_{t|t-1}$

Log-likelihood: $\ell_T(\theta) = -\frac{1}{2} \sum_{t=1}^{T} \left[\log F_t + \frac{v_t^2}{F_t} \right]$

Problem 2: Beveridge–Nelson Trend

Stochastic Trend Decomposition via VAR

Problem 2: Setup

Beveridge–Nelson (1981) trend:

$$\tau_t = x_t + \mathbb{E}_t \sum_{j=1}^{\infty} (\Delta x_{t+j} - \mu_x)$$

Interpretation: The trend τ_t is the level to which x_t is expected to converge after transient dynamics die out.

Setup: Let $y_t = (\Delta x_t, w_t')'$ where:

- Δx_t is the growth of the series of interest
- w_t contains variables that help predict Δx_t

Assume y_t is covariance stationary with mean zero and follows a VAR(2):

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + u_t, \quad u_t \sim \mathcal{N}(0, \Sigma_u)$$

Problem 2(a): Variable Selection

Question: *Explain how you choose variables to include in w_t.*

Solution

Include variables that **Granger-cause** Δx_t.

- A variable w_t Granger-causes Δx_t if past values of w help predict Δx_t beyond what past values of Δx_t alone provide.
- Formally: w_{t-j} appears with nonzero coefficient in the equation for Δx_t.

Practical approach:

- Include leading indicators, related macro variables
- Test for Granger causality; drop variables that don't help
- Balance predictive power vs. parameter proliferation

Problem 2(b): MLE for VAR(2)

Question: *How would you estimate the VAR parameters by MLE?*

Solution: Conditional log-likelihood (given y_0, y_{-1}):

$$\ell_T(\Phi_1, \Phi_2, \Sigma_u) = -\frac{T}{2} \log |\Sigma_u| - \frac{1}{2} \sum_{t=1}^{T} u_t' \Sigma_u^{-1} u_t$$

where $u_t = y_t - \Phi_1 y_{t-1} - \Phi_2 y_{t-2}$.

MLE = OLS Equation-by-Equation

Under Gaussianity, the MLE for (Φ_1, Φ_2) is obtained by OLS on each equation:

$$\hat{\Phi} = \left(\sum_{t=1}^{T} y_t Z_{t-1}' \right) \left(\sum_{t=1}^{T} Z_{t-1} Z_{t-1}' \right)^{-1}$$

where $Z_{t-1} = (y_{t-1}', y_{t-2}')'$ and $\hat{\Phi} = [\hat{\Phi}_1, \hat{\Phi}_2]$.

Problem 2(c): Trend Formula — Companion Form

Question: *Derive an expression for the trend τ_t in terms of the VAR parameters and current and lagged values of y_t.*

Step 1. Write VAR(2) in companion form VAR(1).

- Stack $Y_t = (y_t', y_{t-1}')'$. Then:

$$Y_t = FY_{t-1} + v_t, \quad F = \begin{pmatrix} \Phi_1 & \Phi_2 \\ I & 0 \end{pmatrix}, \quad v_t = \begin{pmatrix} u_t \\ 0 \end{pmatrix}$$

Step 2. Multi-step forecast.

- For stationary VAR(1): $\mathbb{E}_t[Y_{t+h}] = F^h Y_t$.

Step 3. Extract Δx_{t+h} forecast.

- Let $e_1 = (1, 0, \ldots, 0)'$ select Δx_t from Y_t.
- Then: $\mathbb{E}_t[\Delta x_{t+h}] = e_1' F^h Y_t$.

Problem 2(c): Trend Formula — Final Expression

BN trend (definition): $\tau_t = x_t + \sum_{h=1}^{\infty} \mathbb{E}_t[\Delta x_{t+h}]$.

- From Step 3: $\mathbb{E}_t[\Delta x_{t+h}] = e_1' F^h Y_t$, so

$$\tau_t = x_t + \sum_{h=1}^{\infty} e_1' F^h Y_t$$

Geometric series (stable F, eigenvalues inside unit circle):

- $\sum_{h=1}^{\infty} F^h = F(I - F)^{-1}$.

Beveridge–Nelson Trend Formula

$$\boxed{\tau_t = x_t + e_1' F(I - F)^{-1} Y_t}$$

where $Y_t = (y_t', y_{t-1}')'$.

Cycle: $c_t = x_t - \tau_t = -e_1' F(I - F)^{-1} Y_t$.

Problem 2(d): Asymptotic Distribution

Question: *How would you approximate the asymptotic distribution of the coefficients mapping current and lagged values of y_t into τ_t?*

Solution: Let $\psi = \text{vec}(\Phi_1, \Phi_2)$ be the VAR parameters and $h(\psi) = e_1' F(I - F)^{-1}$ be the coefficient mapping Y_t to the trend adjustment.

Delta Method

Since $\sqrt{T}(\hat{\psi} - \psi_0) \xrightarrow{d} \mathcal{N}(0, V_\psi)$, we have:

$$\sqrt{T}(h(\hat{\psi}) - h(\psi_0)) \xrightarrow{d} \mathcal{N}\left(0, \frac{\partial h}{\partial \psi'} V_\psi \frac{\partial h'}{\partial \psi}\right)$$

In practice:

- Compute $\hat{V}_\psi$ from VAR estimation (standard errors)
- Compute Jacobian $\partial h / \partial \psi'$ numerically or analytically
- Or use bootstrap for small-sample inference

Problem 2(e): Random Walk Property

Question: *Prove that τ_t is a random walk (with drift if $\mu_{\Delta x} \neq 0$).*
Proof: From $\tau_t = x_t + e_1' F(I-F)^{-1} Y_t$, take first differences:

$$\begin{aligned}\Delta\tau_t &= \Delta x_t + e_1' F(I-F)^{-1}\Delta Y_t \\ &= \Delta x_t + e_1' F(I-F)^{-1}(Y_t - Y_{t-1})\end{aligned}$$

Using $Y_t = FY_{t-1} + v_t$, we get $\Delta Y_t = (F-I)Y_{t-1} + v_t$:

$$\begin{aligned}\Delta\tau_t &= e_1' v_t + e_1' F(I-F)^{-1}[(F-I)Y_{t-1} + v_t] \\ &= e_1' v_t - e_1' F Y_{t-1} + e_1' F(I-F)^{-1} v_t \\ &= e_1'[I + F(I-F)^{-1}]v_t = e_1'(I-F)^{-1} v_t\end{aligned}$$

Result

$\Delta\tau_t = e_1'(I-F)^{-1} v_t$ is a linear combination of v_t, which is i.i.d.
$\Rightarrow$ τ_t is a **random walk** (with drift if $\mu_{\Delta x} \neq 0$).

Problem 3: Clarida–Galí–Gertler (1999)

NK Model, Policy Regimes, and the Lucas Critique

Problem 3: U.S. Monetary Policy History — Context

Summary of the period (relevant for Burns vs. Volcker and the Lucas critique):

- **1960s–70s**: Great Inflation; Fed often **accommodative** (raised rates less than inflation); Arthur Burns chair (1970–78).
- **1979–87 (Volcker)**: Sharp shift to **anti-inflation** policy; disinflation at cost of recession; Taylor-rule interpretation: ϕ_π increased.
- **Pre-Volcker vs. Volcker**: Clarida, Galí, Gertler (1999) document that estimated Fed reaction to inflation was **weaker** pre-Volcker and **stronger** after; forward-looking Taylor rules fit the regime change.

Recommendation: For a full narrative of U.S. monetary and fiscal policy over 1961–2021 (eight Fed chairs, stagflation to COVID), see Blinder (2022), *A Monetary and Fiscal History of the United States, 1961–2021*, Princeton. **Sources**: Clarida, Galí, Gertler (1999), *J. Econ. Lit.*; Blinder (2022), Princeton University Press.

Problem 3: The NK Model

3-equation NK model:

$$\text{(IS)} \quad x_t = \mathbb{E}_t x_{t+1} - \frac{1}{\sigma}(i_t - \mathbb{E}_t \pi_{t+1} - r^n)$$

$$\text{(NKPC)} \quad \pi_t = \beta \mathbb{E}_t \pi_{t+1} + \kappa x_t + u_t$$

$$\text{(Taylor)} \quad i_t = r^n + \phi_\pi \pi_t$$

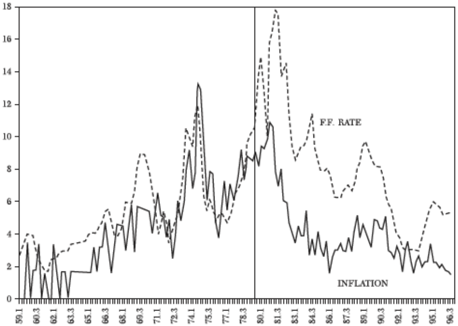
Assumptions:

(A1) $r_t^n = r^n$ constant

(A2) $\mathbb{E}_t x_{t+1} = 0$

(A3) $\mathbb{E}_t \pi_{t+1} = \pi_{t-1}$

(A4) $u_t \sim \text{i.i.d.}(0, \sigma_u^2)$



Federal funds rate and inflation (CGG)

Problem 3(a): Deriving the AR(1) for Inflation — Step 1

Question: Eliminate x_t and i_t; show $\pi_t = a\pi_{t-1} + \varepsilon_t$. Give a, ε_t; sign of $\partial a / \partial \phi_\pi$ and interpret.

- **Step 1**: From IS with (A1)–(A3): $x_t = -\frac{1}{\sigma}(i_t - \pi_{t-1} - r^n)$
- **Step 2**: Substitute Taylor $i_t = r^n + \phi_\pi \pi_t$:

$$x_t = -\frac{1}{\sigma}(\phi_\pi \pi_t - \pi_{t-1})$$

- Hence: $\boxed{x_t = \frac{1}{\sigma}\pi_{t-1} - \frac{\phi_\pi}{\sigma}\pi_t}$

Problem 3(a): Deriving the AR(1) for Inflation — Step 2

Step 3: Substitute x_t into NKPC.

- (A3) $\Rightarrow \mathbb{E}_t\pi_{t+1} = \pi_{t-1}$: NKPC is $\pi_t = \beta\pi_{t-1} + \kappa x_t + u_t$
- Substitute $x_t = \frac{1}{\sigma}\pi_{t-1} - \frac{\phi_\pi}{\sigma}\pi_t$:

$$\pi_t = \beta\pi_{t-1} + \kappa\left(\frac{1}{\sigma}\pi_{t-1} - \frac{\phi_\pi}{\sigma}\pi_t\right) + u_t$$

- Collect π_t and π_{t-1}:

$$\pi_t\left(1 + \frac{\kappa\phi_\pi}{\sigma}\right) = \left(\beta + \frac{\kappa}{\sigma}\right)\pi_{t-1} + u_t$$

Problem 3(a): AR(1) Result

Reduced-Form AR(1) for Inflation

$\pi_t = a \cdot \pi_{t-1} + \varepsilon_t$ with $a = \frac{\beta\sigma+\kappa}{\sigma+\kappa\phi_\pi}$, $\varepsilon_t = \dfrac{\sigma u_t}{\sigma + \kappa\phi_\pi}$

- **Sign**: $\dfrac{\partial a}{\partial \phi_\pi} = \dfrac{-\kappa(\beta\sigma + \kappa)/\sigma}{(\sigma + \kappa\phi_\pi)^2} < 0$
- **Interpretation**: More aggressive policy ($\uparrow \phi_\pi$) $\Rightarrow$ lower persistence ($\downarrow a$)

Problem 3(b): Burns vs. Volcker Hypothesis

Question: Translate "policy changed from Burns to Volcker" into H_0 and H_1 in terms of a_1, a_2.

- Break at T_b (1979:Q3): $\phi_\pi = \phi_{\pi,1}$ for $t \leq T_b$, $\phi_\pi = \phi_{\pi,2}$ for $t > T_b$
- $\partial a / \partial \phi_\pi < 0 \Rightarrow$ Burns $\rightarrow a_1$, Volcker $\rightarrow a_2$
- "Volcker more aggressive" $\Rightarrow \phi_{\pi,2} > \phi_{\pi,1} \Rightarrow a_2 < a_1$

Hypotheses

$H_0 : a_1 = a_2$ (no change) vs. $H_1 : a_1 > a_2$ (Volcker more aggressive)

Problem 3(c): Conditional Log-Likelihood

Question: (i) Conditional log-likelihood $\ell_n(\theta)$ given π_0, $\theta = (c, a_1, a_2, \sigma^2)$. (ii) LR test for $H_0 : a_1 = a_2$ and its asymptotic distribution.

$$t \leq T_b: \quad \pi_t = c + a_1 \pi_{t-1} + \varepsilon_t$$
$$t > T_b: \quad \pi_t = c + a_2 \pi_{t-1} + \varepsilon_t$$
$$\varepsilon_t \mid \mathcal{F}_{t-1} \sim \mathcal{N}(0, \sigma^2)$$

Conditional Log-Likelihood (given π_0)

The sum **splits at the break**: use a_1 for $t \leq T_b$, a_2 for $t > T_b$:

$$\ell_n(\theta) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \left[\sum_{t=1}^{T_b} (\pi_t - c - a_1 \pi_{t-1})^2 + \sum_{t=T_b+1}^{n} (\pi_t - c - a_2 \pi_{t-1})^2 \right]$$

Quick review: Asymptotic tests

Where do we use them?

- As $n \to \infty$, asymptotic approximation works better; we can derive tests even in complicated problems where no optimal (finite-sample) test exists.

Trinity of large-sample tests

- **LR**: distance between log-likelihoods (restricted vs. unrestricted)
- **Wald**: distance between estimators ($\hat{\theta}$ vs. value under H_0)
- **Score (LM)**: distance to zero score (score at restricted MLE)

$$S(\theta) = \nabla_\theta \ell_n(\theta) = \left(\tfrac{\partial \ell_n}{\partial \theta_1}, \ldots, \tfrac{\partial \ell_n}{\partial \theta_k}\right)'$$

Differences (what to estimate)

- **LR**: estimate both restricted and unrestricted models
- **Wald**: estimate only unrestricted model (for simple null)
- **LM**: estimate only restricted model

Problem 3(c): LR Test for $H_0 : a_1 = a_2$

- **Unrestricted**: MLE $\hat{\theta}_U = (\hat{c}, \hat{a}_1, \hat{a}_2, \hat{\sigma}^2)$

$$\mathrm{RSS}_U = \sum_{t=1}^{T_b} (\pi_t - \hat{c} - \hat{a}_1 \pi_{t-1})^2 + \sum_{t=T_b+1}^{n} (\pi_t - \hat{c} - \hat{a}_2 \pi_{t-1})^2$$

- **Restricted**: impose $a_1 = a_2 = a$, MLE $\tilde{\theta}_R = (\tilde{c}, \tilde{a}, \tilde{\sigma}^2)$

$$\mathrm{RSS}_R = \sum_{t=1}^{n} (\pi_t - \tilde{c} - \tilde{a}\pi_{t-1})^2$$

- **MLE variance**: $\hat{\sigma}^2 = \mathrm{RSS}/n$ for each model
- **Statistic**: $LR = 2[\ell_n(\hat{\theta}_U) - \ell_n(\hat{\theta}_R)]$
- **Derivation**: plug $\hat{\sigma}^2 = \mathrm{RSS}/n$ into ℓ_n, simplify $\Rightarrow$

$$LR = n \log\left(\frac{\mathrm{RSS}_R}{\mathrm{RSS}_U}\right) \geq 0$$

- **Under** H_0: $LR \xrightarrow{d} \chi^2(1)$ (one restriction: $a_1 - a_2 = 0$) ▸ General LR theory

Problem 3(d): Wald Test for $H_0 : a_1 = a_2$

Wald: uses **unrestricted** MLE $\hat{\theta}_U = (\hat{c}, \hat{a}_1, \hat{a}_2, \hat{\sigma}^2)$

- **Restriction**: $r(\theta) = a_1 - a_2 = 0$
- **Statistic**: $W = \dfrac{(\hat{a}_1 - \hat{a}_2)^2}{\widehat{\text{Var}}(\hat{a}_1 - \hat{a}_2)}$
- **Denominator**: $\widehat{\text{Var}}(\hat{a}_1 - \hat{a}_2)$ comes from the **estimated asymptotic covariance matrix** of $\hat{\theta}_U$ (e.g. inverse Hessian or OLS $(X'X)^{-1}$ block).
- **Under** H_0: $W \xrightarrow{d} \chi^2(1)$

▸ See Appendix: General Wald theory

Problem 3(d): LM (Score) Test for $H_0 : a_1 = a_2$

Restricted MLE $\tilde{\theta}_R = (\tilde{c}, \tilde{a}, \tilde{\sigma}^2)$; $\tilde{\varepsilon}_t = \pi_t - \tilde{c} - \tilde{a}\pi_{t-1}$.

- **Joint score** (sum over $t = 1, \dots, n$; same $\tilde{\varepsilon}_t$ under restricted model): $S_{a_1} = \frac{1}{\tilde{\sigma}^2} \sum_{t=1}^{T_b} \tilde{\varepsilon}_t \pi_{t-1}$, $\quad S_{a_2} = \frac{1}{\tilde{\sigma}^2} \sum_{t=T_b+1}^{n} \tilde{\varepsilon}_t \pi_{t-1}$. Restricted FOC $\Rightarrow S_{a_1} + S_{a_2} = 0$.
- **Hypothesis only on a's** $\Rightarrow$ use score $S_a = (S_{a_1}, S_{a_2})'$ and the (a_1, a_2) **block** of the Fisher information.
- **Fisher information for** (a_1, a_2): $\frac{\partial^2 \ell}{\partial a_1 \partial a_2} = 0$ (disjoint sums) $\Rightarrow$

$$\mathcal{I}_{aa} = \frac{1}{\tilde{\sigma}^2} \begin{bmatrix} q_1 & 0 \\ 0 & q_2 \end{bmatrix}, \quad q_1 = \sum_{t=1}^{T_b} \pi_{t-1}^2,\ q_2 = \sum_{t=T_b+1}^{n} \pi_{t-1}^2$$

  (When c is also estimated, use the (a_1, a_2) block of the *inverse* of the full (c, a_1, a_2) information.)
- **LM**: $LM = S_a' \, \mathcal{I}_{aa}^{-1} \, S_a$ (or with the full-inverse (a_1, a_2) block). Under H_0: $LM \xrightarrow{d} \chi^2(1)$.

See Appendix: General LM theory

Problem 3(e): Interpretation

Question: If $\hat{a}_2 < \hat{a}_1$ and we reject H_0, interpret $\phi_{\pi,2}$ vs. $\phi_{\pi,1}$.

- Persistence **fell** after T_b
- $\partial a / \partial \phi_\pi < 0 \Rightarrow \phi_{\pi,2} > \phi_{\pi,1}$
- **Volcker more aggressive** than Burns

Problem 3(f): Lucas Critique

Question: Why does an AR(1) estimated under Burns fail to forecast under Volcker? Structural vs. reduced-form.

Why AR(1) from Burns Fails Under Volcker

- **Structural**: ϕ_π (policy), κ, β, σ (deep)
- **Reduced-form**: $a = \frac{\beta\sigma+\kappa}{\sigma+\kappa\phi_\pi}$ (depends on policy)
- Burns-era AR(1) uses $a_1 = a(\phi_{\pi,1})$; after switch, true $a = a_2 \neq a_1$
- $\Rightarrow$ **Forecasts based on a_1 systematically wrong**

Problem 4: Kalman Filter

With Correlated State and Measurement Innovations

Problem 4: Question

Question: Derive the Kalman filter for this model.

$$S_t = AS_{t-1} + B\varepsilon_{1t} \qquad \text{(State)}$$

$$X_t = CS_t + D\varepsilon_{2t} \qquad \text{(Measurement)}$$

Shocks are normally distributed:

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} Q & F \\ F' & R \end{pmatrix}\right)$$

Innovation: $F \neq 0 \quad \Rightarrow \quad$ measurement shock correlated with structural (state) shock.

Problem 4: Definitions

- **Predicted state:** $\hat{S}_{t|t-1} = \mathbb{E}[S_t \mid X_{t-1}, \dots, X_1]$
- **Prediction MSE:** $P_{t|t-1} = \mathrm{Var}(S_t \mid X_{t-1}, \dots, X_1)$
- **Innovation (prediction error):** $\nu_t = X_t - C\hat{S}_{t|t-1}$
- **Innovation variance:** $\Sigma_\nu = \mathrm{Var}(\nu_t)$
- **Kalman gain:** K_t (weight on ν_t in the update)
- **Updated state:** $\hat{S}_{t|t} = \mathbb{E}[S_t \mid X_t, \dots, X_1]$, $P_{t|t} = \mathrm{Var}(S_t \mid X_t, \dots, X_1)$

Problem 4: Modified Kalman Filter — Algorithm

Initialization: $\hat{S}_{0|0}$, $P_{0|0}$ (e.g. unconditional or prior). **For** $t = 1, \ldots, T$: *Prediction step* (same as standard; no F):

$$\hat{S}_{t|t-1} = A\hat{S}_{t-1|t-1}, \qquad P_{t|t-1} = AP_{t-1|t-1}A' + BQB'$$

Update step (upon observing X_t):

1. **Prediction error (innovation)**: $\nu_t = X_t - C\hat{S}_{t|t-1}$
2. **Innovation variance** (modified; cross terms in F): ▸ Derivation

$$\Sigma_\nu = CP_{t|t-1}C' + DRD' + CBFD' + DF'B'C'$$

3. **Kalman gain** (modified): $K_t = (P_{t|t-1}C' + BFD')\,\Sigma_\nu^{-1}$
4. **Updated state estimate**: $\hat{S}_{t|t} = \hat{S}_{t|t-1} + K_t\nu_t$
5. **Updated state MSE**: $P_{t|t} = P_{t|t-1} - K_t\Sigma_\nu K_t'$

Problem 5: ARMA(2,2) to State Space

Converting ARMA to State-Space Form

Problem 5: Setup

Question: Put ARMA(2,2) into state-space form and verify.

- **ARMA(2,2)**: $(1 - \rho_1 L - \rho_2 L^2) y_t = (1 + \theta_1 L + \theta_2 L^2)\varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2)$
- **Expanded**: $y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$
- **Goal**: State-space $S_t = AS_{t-1} + Bu_t, \quad y_t = CS_t$

Problem 5: State-Space Representation

- **State**: $S_t = (y_t, y_{t-1}, \varepsilon_t, \varepsilon_{t-1})'$
- **State equation** $S_t = AS_{t-1} + B\varepsilon_t$:

$$\begin{pmatrix} y_t \\ y_{t-1} \\ \varepsilon_t \\ \varepsilon_{t-1} \end{pmatrix} = \begin{pmatrix} \rho_1 & \rho_2 & \theta_1 & \theta_2 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \varepsilon_{t-1} \\ \varepsilon_{t-2} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \varepsilon_t$$

- **Measurement**: $y_t = CS_t, \quad C = (1, 0, 0, 0)$

Problem 5: Verification

- **Verify**: First row of state equation gives

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_t$$

- Matches ARMA(2,2): $y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$ ✓

Summary

State $S_t = (y_t, y_{t-1}, \varepsilon_t, \varepsilon_{t-1})'$; A (4×4 above), $B = (1, 0, 1, 0)'$; $C = (1, 0, 0, 0)$, $D = 0$

Summary

Problem 1 (Bansal–Yaron):

- VAR(1) representation; stationarity requires $|\rho| < 1$
- Kalman filter for MLE when state is latent

Problem 2 (Beveridge–Nelson):

- Trend formula: $\tau_t = x_t + e_1' F(I - F)^{-1} Y_t$; BN trend is a random walk

Problem 3 (CGG):

- AR(1) persistence $a = \frac{\beta\sigma+\kappa}{\sigma+\kappa\phi_\pi}$; $\partial a/\partial\phi_\pi < 0$
- LR, Wald, LM tests for structural break; Lucas critique

Problem 4: Modified Kalman gain when innovations correlated
Problem 5: ARMA(2,2) $\rightarrow$ 4-dimensional state space

Appendix

Appendix: General LR Test Theory

LR test (one of the most useful methods for complicated problems):

- Reject H_0 if $x \in \left\{ x : \lambda(x) = \frac{\sup_{\theta \in \Theta_0} \ell(\theta|x)}{\sup_{\theta \in \Theta} \ell(\theta|x)} \leq c \right\}$
- Even if we cannot obtain the suprema analytically, we can compute them numerically
- Choose c so that $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\lambda(X) \leq c) \leq \alpha$ (α is the the level of the test)

Asymptotic distribution (under regularity conditions):

- Under $H_0 : \theta = \theta_0$: $\quad -2 \ln \lambda(X) \xrightarrow{d} \chi^2_q \quad (q = \#\text{ restrictions})$
- **Proof sketch**: Taylor expand $\ln \ell(\theta|x)$ around $\hat{\theta}$:

$$-2 \ln \lambda(x) = 2[\ln \ell(\hat{\theta}|x) - \ln \ell(\theta_0|x)] \approx -\ln \ell''(\theta_0|x)(\theta_0 - \hat{\theta})^2$$

- Use $-\frac{1}{n} \ln \ell''(\hat{\theta}|x) \xrightarrow{p} \mathcal{I}(\theta_0)$ and $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \mathcal{I}(\theta_0)^{-1})$

Caveat: The above is for i.i.d. samples. With time series, replace with appropriate regularity conditions (e.g. ergodicity, mixing) for dependent data.

Appendix: LR Derivation for Gaussian Regression

Setup: Gaussian log-likelihood

$$\ell_n(\theta) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\text{RSS}(\theta)$$

- **MLE for σ^2**: $\hat{\sigma}^2 = \text{RSS}/n$

$$\ell_n(\hat{\theta}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\left(\frac{\text{RSS}}{n}\right) - \frac{n}{2}$$

- **LR statistic**:

$$\begin{aligned} LR &= 2[\ell_n(\hat{\theta}_U) - \ell_n(\hat{\theta}_R)] = 2\left[-\frac{n}{2}\log\left(\frac{\text{RSS}_U}{n}\right) + \frac{n}{2}\log\left(\frac{\text{RSS}_R}{n}\right)\right] \\ &= n\log\left(\frac{\text{RSS}_R}{\text{RSS}_U}\right) \end{aligned}$$

Result

$LR = n\log(\text{RSS}_R/\text{RSS}_U) \xrightarrow{d} \chi^2(q)$ under H_0 ($q = \#$ restrictions)

Appendix: General Wald Test Theory

Wald test: tests $H_0 : r(\theta) = 0$ using **unrestricted** MLE $\hat{\theta}$

- **Idea**: If H_0 is true, $r(\hat{\theta})$ should be close to zero.
- **Statistic**:
$$W = r(\hat{\theta})' \left[R(\hat{\theta}) \, \widehat{\text{Var}}(\hat{\theta}) \, R(\hat{\theta})' \right]^{-1} r(\hat{\theta})$$
where $R(\theta) = \partial r / \partial \theta'$ is the Jacobian of restrictions.
- **Scalar case** ($q = 1$ restriction): $W = \dfrac{[r(\hat{\theta})]^2}{\widehat{\text{Var}}(r(\hat{\theta}))}$
- **Under** H_0: $W \xrightarrow{d} \chi^2(q)$ ($q = \#$ restrictions)

Advantage: Only need to estimate the unrestricted model.

Caveat: For time series, use HAC standard errors if innovations are serially correlated.

Appendix: General LM (Score) Test Theory

LM test: tests $H_0 : r(\theta) = 0$ using **restricted** MLE $\tilde{\theta}$

- **Idea**: Under H_0, the score $S(\theta_0) = \nabla_\theta \ell_n(\theta_0)$ has mean zero. If $S(\tilde{\theta})$ is "large," reject H_0.
- **Statistic**:
$$LM = S(\tilde{\theta})' \mathcal{I}(\tilde{\theta})^{-1} S(\tilde{\theta})$$
where $\mathcal{I}(\theta) = -\mathbb{E}[\nabla_\theta^2 \ell_n(\theta)]$ is the Fisher information.
- **Under** H_0: $LM \xrightarrow{d} \chi^2(q)$ ($q = \#$ restrictions)

Advantage: Only need to estimate the restricted model (useful when unrestricted is hard to estimate).
Caveat: For time series, replace Fisher information with appropriate long-run variance if needed.

Appendix: Trinity — Asymptotic Equivalence

Under regularity conditions and H_0:

$$LR,\ W,\ LM \xrightarrow{d} \chi^2(q)$$

- All three are **asymptotically equivalent** under the null.
- In finite samples, they can differ; no universal ranking.
- **Rule of thumb**:
  - Use **LR** if you can estimate both models easily.
  - Use **Wald** if unrestricted model is easy, restricted is hard.
  - Use **LM** if restricted model is easy, unrestricted is hard.

Time series caveat: The i.i.d. theory extends to dependent data under ergodicity/mixing conditions; may need HAC variance estimators.

Appendix: Innovation Variance Derivation

Goal: Derive $\Sigma_\nu = \text{Var}(\nu_t)$ when $\text{Cov}(\varepsilon_{1t}, \varepsilon_{2t}) = F \neq 0$.

Step 1: Write the innovation in terms of prediction error $\tilde{S}_t = S_t - \hat{S}_{t|t-1}$:

$$\nu_t = X_t - C\hat{S}_{t|t-1} = C(S_t - \hat{S}_{t|t-1}) + D\varepsilon_{2t} = C\tilde{S}_t + D\varepsilon_{2t}$$

Step 2: Compute variance (key: $\tilde{S}_t$ involves ε_{1t}, and $\text{Cov}(\varepsilon_{1t}, \varepsilon_{2t}) = F$):

$$\begin{aligned}\Sigma_\nu &= \text{Var}(C\tilde{S}_t + D\varepsilon_{2t}) \\ &= C\underbrace{\text{Var}(\tilde{S}_t)}_{P_{t|t-1}}C' + D\underbrace{\text{Var}(\varepsilon_{2t})}_{R}D' + C\text{Cov}(\tilde{S}_t, \varepsilon_{2t})D' + D\text{Cov}(\varepsilon_{2t}, \tilde{S}_t)C'\end{aligned}$$

Since $S_t = AS_{t-1} + B\varepsilon_{1t}$ with $S_{t-1} \perp \varepsilon_{2t}$, and $\hat{S}_{t|t-1} \perp \varepsilon_{2t}$:
$\text{Cov}(\tilde{S}_t, \varepsilon_{2t}) = B\text{Cov}(\varepsilon_{1t}, \varepsilon_{2t}) = BF$.

$$\boxed{\Sigma_\nu = CP_{t|t-1}C' + DRD' + CBFD' + DF'B'C'}$$

Appendix: Bayesian Updating in Linear Gaussian Models

Key fact: If (X, Y) are jointly normal, then $X \mid Y$ is also normal with:

$$\mathbb{E}[X \mid Y] = \mathbb{E}[X] + \text{Cov}(X, Y)\,\text{Var}(Y)^{-1}(Y - \mathbb{E}[Y])$$

$$\text{Var}(X \mid Y) = \text{Var}(X) - \text{Cov}(X, Y)\,\text{Var}(Y)^{-1}\text{Cov}(Y, X)$$

Application to Kalman filter: Let $X = S_t$ (state), $Y = \nu_t$ (innovation). Then:

- $\text{Cov}(S_t, \nu_t) = \text{Cov}(\tilde{S}_t, \nu_t) = P_{t|t-1}C' + BFD'$
- $\text{Var}(\nu_t) = \Sigma_\nu$

The Kalman update is exactly Bayesian updating:

$$K_t = \text{Cov}(\tilde{S}_t, \nu_t)\,\Sigma_\nu^{-1} = (P_{t|t-1}C' + BFD')\,\Sigma_\nu^{-1}$$

$$\hat{S}_{t|t} = \hat{S}_{t|t-1} + K_t\nu_t \quad \text{(conditional mean)}$$

$$P_{t|t} = P_{t|t-1} - K_t\Sigma_\nu K_t' \quad \text{(conditional variance)}$$

Appendix: Kalman Gain Derivation

Goal: Derive $K_t = (P_{t|t-1}C' + BFD')\Sigma_\nu^{-1}$.

Step 1: From Bayesian updating, $K_t = \text{Cov}(\tilde{S}_t, \nu_t)\,\Sigma_\nu^{-1}$.

Step 2: Compute $\text{Cov}(\tilde{S}_t, \nu_t)$ where $\nu_t = C\tilde{S}_t + D\varepsilon_{2t}$:

$$\begin{aligned}
\text{Cov}(\tilde{S}_t, \nu_t) &= \text{Cov}(\tilde{S}_t, C\tilde{S}_t + D\varepsilon_{2t}) \\
&= \text{Cov}(\tilde{S}_t, C\tilde{S}_t) + \text{Cov}(\tilde{S}_t, D\varepsilon_{2t}) \\
&= \text{Var}(\tilde{S}_t)C' + \text{Cov}(\tilde{S}_t, \varepsilon_{2t})D' \\
&= P_{t|t-1}C' + BF \cdot D'
\end{aligned}$$

(using $\text{Cov}(\tilde{S}_t, \varepsilon_{2t}) = BF$ from the innovation variance derivation)

Step 3: Substitute into Bayesian formula:

$$\boxed{K_t = (P_{t|t-1}C' + BFD')\,\Sigma_\nu^{-1}}$$

Note: When $F = 0$, this reduces to standard Kalman gain $K_t = P_{t|t-1}C'\Sigma_\nu^{-1}$.