**EI035 Econometrics II**
The Graduate Institute, Spring 2025, Marko Mlikota

# PS2 Solutions

**Jingle Fu**

# Problem 1: Dynamic Panel Data with Correlated Random Effects

## Model

$$y_{it} = \alpha_i + \rho y_{it-1} + u_{it}, \quad u_{it} \sim iid\mathcal{N}(0,1)$$

## CRE Distribution

$$\alpha_i | (y_{i0}, \phi) \sim \mathcal{N}(\phi y_{i0}, 1)$$

## (a) The Incidental Parameter Problem (IPP)

The incidental parameter problem arises in panel data models when the number of parameters to be estimated grows with the sample size $N$. Here, the unit-specific effects $\alpha_1, ..., \alpha_N$ are the incidental parameters.
In a dynamic panel (where $y_{it-1}$ is a regressor), the standard Fixed Effects (Within) estimator or the naive MLE for $\alpha_i$ and $\rho$ yields inconsistent estimates for $\rho$ when $N \to \infty$ but $T$ remains fixed.
Treating $\{\alpha_i\}_{i=1}^n$ as fixed parameters in FE–ML yields

$$\ell_i(\alpha_i, \rho) \;=\; -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\big(y_{it} - \alpha_i - \rho\, y_{i,t-1}\big)^2,$$

and the first-order condition for $\alpha_i$ is

$$\frac{\partial \ell_i}{\partial \alpha_i} \;=\; \sum_{t=1}^{T}\big(y_{it} - \alpha_i - \rho\, y_{i,t-1}\big) \;=\; 0 \;=\; T\,\hat{\alpha}_i - \sum_{t=1}^{T}\big(y_{it} - \rho\, y_{i,t-1}\big),$$

hence

$$\hat{\alpha}_i \;=\; \frac{1}{T}\sum_{t=1}^{T}\big(y_{it} - \rho\, y_{i,t-1}\big) \;=\; \alpha_i \;+\; \frac{1}{T}\sum_{t=1}^{T} u_{it}.$$

Because $y_{i,t-1}$ embeds $\alpha_i$, the estimation error $\hat{\alpha}_i - \alpha_i$ remains correlated with $y_{i,t-1}$ when $T$ is fixed:

$$\mathrm{Cov}\big(y_{i,t-1},\ \hat{\alpha}_i - \alpha_i\big) \ \neq\ 0 \quad \text{for fixed } T.$$

Thus, as $n \to \infty$ with $T$ fixed, the FE estimator $\hat{\rho}$ has an $O(1/T)$ Nickell bias. In contrast, the CRE–Bayesian route treats $\alpha_i$ hierarchically so that inference on $(\phi, \rho)$ does not suffer from the IPP.

## (b) Integrating out $\alpha_i$

Given $\tilde{y}_{it} = \alpha_i + u_{it}$ and $u_{it} \sim \mathcal{N}(0,1)$, we have:

$$p(\tilde{y}_{it} \mid \alpha_i) = \frac{1}{\sqrt{2\pi}} \exp\Big( -\tfrac{1}{2}(\tilde{y}_{it} - \alpha_i)^2 \Big).$$

Due to independence across $t$ (conditional on $\alpha_i$):

$$\begin{aligned} p(\tilde{\mathbf{y}}_i \mid \alpha_i) &= \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi}} \exp\Big( -\tfrac{1}{2}(\tilde{y}_{it} - \alpha_i)^2 \Big) \\ &= (2\pi)^{-T/2} \exp\Big( -\tfrac{1}{2} \sum_{t=1}^{T} (\tilde{y}_{it} - \alpha_i)^2 \Big). \end{aligned}$$

The prior (CRE) is:

$$p(\alpha_i \mid y_{i0}, \phi) = \frac{1}{\sqrt{2\pi}} \exp\Big( -\tfrac{1}{2}(\alpha_i - \phi y_{i0})^2 \Big).$$

Thus:

$$\begin{aligned} p(\tilde{\mathbf{y}}_i, \alpha_i \mid y_{i0}, \phi, \rho) &= p(\tilde{\mathbf{y}}_i \mid \alpha_i)\, p(\alpha_i \mid y_{i0}, \phi) \\ &= (2\pi)^{-(T+1)/2} \exp\Big( -\tfrac{1}{2} \underbrace{\Big[ \sum_{t=1}^{T} (\tilde{y}_{it} - \alpha_i)^2 + (\alpha_i - \phi y_{i0})^2 \Big]}_{\mathcal{Q}(\alpha_i)} \Big). \end{aligned}$$

$$\begin{aligned} \sum_{t=1}^{T} (\tilde{y}_{it} - \alpha_i)^2 &= \sum_{t=1}^{T} \tilde{y}_{it}^2 - 2\alpha_i \sum_{t=1}^{T} \tilde{y}_{it} + T\alpha_i^2, \\ (\alpha_i - \phi y_{i0})^2 &= \alpha_i^2 - 2\phi y_{i0}\alpha_i + \phi^2 y_{i0}^2. \end{aligned}$$

Summing these yields:

$$\mathcal{Q}(\alpha_i) = (T+1)\alpha_i^2 - 2\alpha_i \Big( \sum_{t=1}^{T} \tilde{y}_{it} + \phi y_{i0} \Big) + \Big( \sum_{t=1}^{T} \tilde{y}_{it}^2 + \phi^2 y_{i0}^2 \Big).$$

Let:

$$c \equiv T+1, \quad b \equiv \sum_{t=1}^{T} \tilde{y}_{it} + \phi y_{i0}, \quad a \equiv \sum_{t=1}^{T} \tilde{y}_{it}^2 + \phi^2 y_{i0}^2,$$

then:

$$\mathcal{Q}(\alpha_i) = c\alpha_i^2 - 2b\alpha_i + a = c\Big(\alpha_i - \frac{b}{c}\Big)^2 + \Big(a - \frac{b^2}{c}\Big).$$

$$\begin{aligned} p(\tilde{\mathbf{y}}_i \mid y_{i0}, \phi, \rho) &= \int_{-\infty}^{\infty} p(\tilde{\mathbf{y}}_i, \alpha_i \mid \cdot)\, d\alpha_i \\ &= (2\pi)^{-(T+1)/2} \exp\Big( -\tfrac{1}{2}\Big(a - \tfrac{b^2}{c}\Big)\Big) \int_{-\infty}^{\infty} \exp\Big( -\tfrac{1}{2}c\big(\alpha_i - \tfrac{b}{c}\big)^2\Big)\, d\alpha_i \\ &= (2\pi)^{-(T+1)/2} \exp\Big( -\tfrac{1}{2}\Big(a - \tfrac{b^2}{c}\Big)\Big) \cdot \sqrt{\frac{2\pi}{c}} \\ &= (2\pi)^{-T/2}\, c^{-1/2}\, \exp\Big( -\tfrac{1}{2}\Big[a - \tfrac{b^2}{c}\Big]\Big). \end{aligned}$$

Substituting $c = T+1$ back:

$$p(\tilde{\mathbf{y}}_i \mid y_{i0}, \phi, \rho) = (2\pi)^{-T/2}(T+1)^{-1/2} \exp\Big( -\tfrac{1}{2}\Big[\sum_{t=1}^{T} \tilde{y}_{it}^2 + \phi^2 y_{i0}^2 - \frac{\big(\sum_{t=1}^{T} \tilde{y}_{it} + \phi y_{i0}\big)^2}{T+1}\Big]\Big).$$

Note that:

$$\sum_{t=1}^{T} \tilde{y}_{it}^2 = \tilde{\mathbf{y}}_i^\top \tilde{\mathbf{y}}_i, \qquad \sum_{t=1}^{T} \tilde{y}_{it} = \mathbf{1}^\top \tilde{\mathbf{y}}_i.$$

Let $\boldsymbol{\mu} \equiv \phi y_{i0} \mathbf{1}$ and

$$\boldsymbol{\Omega} \equiv I_T + \mathbf{1}\mathbf{1}^\top \Rightarrow \boldsymbol{\Omega}^{-1} = I_T - \frac{1}{T+1}\mathbf{1}\mathbf{1}^\top, \quad |\boldsymbol{\Omega}| = (1+T)\cdot 1^{T-1} = T+1.$$

$$
\begin{aligned}
\mathcal{Q} &= \sum_{t=1}^{T} \tilde{y}_{it}^2 + \phi^2 y_{i0}^2 - \frac{\left(\sum_{t=1}^{T} \tilde{y}_{it} + \phi y_{i0}\right)^2}{T+1} \\
&= \tilde{\mathbf{y}}_i^\top \tilde{\mathbf{y}}_i - \frac{(\sum \tilde{y}_{it})^2}{T+1} - \frac{2\phi y_{i0}}{T+1} \sum \tilde{y}_{it} + \phi^2 y_{i0}^2 \Big(1 - \frac{1}{T+1}\Big) \\
&= \tilde{\mathbf{y}}_i^\top \tilde{\mathbf{y}}_i - \frac{(\sum \tilde{y}_{it})^2}{T+1} + \Big[(\phi y_{i0})^2 T - \frac{(\phi y_{i0})^2 T^2}{T+1}\Big] + \Big[-2\phi y_{i0} \sum \tilde{y}_{it} + \frac{2\phi y_{i0} T}{T+1} \sum \tilde{y}_{it}\Big] \\
&= \tilde{\mathbf{y}}_i^\top \tilde{\mathbf{y}}_i + (\phi y_{i0})^2 T - 2\phi y_{i0} \sum \tilde{y}_{it} - \frac{1}{T+1}\Big[(\sum \tilde{y}_{it})^2 - 2\phi y_{i0} T \sum \tilde{y}_{it} + (\phi y_{i0})^2 T^2\Big] \\
&= \tilde{\mathbf{y}}_i^\top \tilde{\mathbf{y}}_i - 2(\phi y_{i0}) \mathbf{1}^\top \tilde{\mathbf{y}}_i + (\phi y_{i0})^2 \mathbf{1}^\top \mathbf{1} - \frac{1}{T+1}\big(\mathbf{1}^\top \tilde{\mathbf{y}}_i - \phi y_{i0} \mathbf{1}^\top \mathbf{1}\big)^2 \\
&= \tilde{\mathbf{y}}_i^\top \tilde{\mathbf{y}}_i - 2\boldsymbol{\mu}^\top \tilde{\mathbf{y}}_i + \boldsymbol{\mu}^\top \boldsymbol{\mu} - \frac{1}{T+1}\big(\mathbf{1}^\top \tilde{\mathbf{y}}_i - \mathbf{1}^\top \boldsymbol{\mu}\big)^2 \\
&= (\tilde{\mathbf{y}}_i - \boldsymbol{\mu})^\top (\tilde{\mathbf{y}}_i - \boldsymbol{\mu}) - \frac{1}{T+1}\big[\mathbf{1}^\top (\tilde{\mathbf{y}}_i - \boldsymbol{\mu})\big]^2 \\
&= (\tilde{\mathbf{y}}_i - \boldsymbol{\mu})^\top \Big(I_T - \tfrac{1}{T+1} \mathbf{1}\mathbf{1}^\top\Big) (\tilde{\mathbf{y}}_i - \boldsymbol{\mu}) \\
&= (\tilde{\mathbf{y}}_i - \boldsymbol{\mu})^\top \boldsymbol{\Omega}^{-1} (\tilde{\mathbf{y}}_i - \boldsymbol{\mu})
\end{aligned}
$$

Therefore:

$$
p(\tilde{\mathbf{y}}_i \mid y_{i0}, \phi, \rho) = (2\pi)^{-T/2} |\boldsymbol{\Omega}|^{-1/2} \exp\Big(-\tfrac{1}{2}(\tilde{\mathbf{y}}_i - \boldsymbol{\mu})^\top \boldsymbol{\Omega}^{-1} (\tilde{\mathbf{y}}_i - \boldsymbol{\mu})\Big).
$$

## (c) Consistency of $(\phi, \rho)$

The sample log-likelihood is:

$$
\ell(\phi, \rho) = \sum_{i=1}^{n} \ell_i(\phi, \rho),
$$

where (derived directly from the above equation):

$$
\ell_i = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \log(T+1) - \frac{1}{2} (\tilde{\mathbf{y}}_i - \phi y_{i0} \mathbf{1})^\top \Big(I_T - \frac{1}{T+1} \mathbf{1}\mathbf{1}^\top\Big) (\tilde{\mathbf{y}}_i - \phi y_{i0} \mathbf{1}).
$$

**Score w.r.t. $\phi$:**

$$
\begin{aligned}
\frac{\partial \ell}{\partial \phi} &= \sum_{i=1}^{n} \frac{\partial \ell_i}{\partial \phi} = \sum_{i=1}^{n} \frac{1}{2} \cdot 2\, y_{i0} \mathbf{1}^\top \Big(I - \tfrac{1}{T+1} \mathbf{1}\mathbf{1}^\top\Big) (\tilde{\mathbf{y}}_i - \phi y_{i0} \mathbf{1}) \cdot (+1) \\
&= \sum_{i=1}^{n} y_{i0}\, \mathbf{1}^\top \Big(I - \tfrac{1}{T+1} \mathbf{1}\mathbf{1}^\top\Big) (\tilde{\mathbf{y}}_i - \phi y_{i0} \mathbf{1}).
\end{aligned}
$$

Setting this score to 0 implies:

$$
\sum_i y_{i0}\, \mathbf{1}^\top \Big(I - \tfrac{1}{T+1} \mathbf{1}\mathbf{1}^\top\Big) \tilde{\mathbf{y}}_i = \phi \sum_i y_{i0}^2\, \mathbf{1}^\top \Big(I - \tfrac{1}{T+1} \mathbf{1}\mathbf{1}^\top\Big) \mathbf{1}.
$$

In the RHS, $\mathbf{1}^\top(I - \frac{1}{T+1}\mathbf{1}\mathbf{1}^\top)\mathbf{1} = \mathbf{1}^\top\mathbf{1} - \frac{1}{T+1}\mathbf{1}^\top\mathbf{1}\,\mathbf{1}^\top\mathbf{1} = T - \frac{T^2}{T+1} = \frac{T}{T+1}$.

**Score w.r.t. $\rho$:**

$$\tilde{\mathbf{y}}_i = \mathbf{y}_i - \rho\,\mathbf{L}_i\mathbf{y}_i \quad \Rightarrow \quad \frac{\partial \tilde{\mathbf{y}}_i}{\partial \rho} = -(\mathbf{L}_i\mathbf{y}_i).$$

Thus:

$$\frac{\partial \ell}{\partial \rho} = \sum_i \big[-(\mathbf{L}_i\mathbf{y}_i)\big]^\top \Big(I - \tfrac{1}{T+1}\mathbf{1}\mathbf{1}^\top\Big)(\tilde{\mathbf{y}}_i - \phi y_{i0}\mathbf{1}) = 0.$$

These two **moment equations** have an **expectation** of 0 at the true values $(\phi_0, \rho_0)$. As $n \to \infty$ with fixed $T$, the sample average score converges to its expectation, and provided the information matrix is non-degenerate (positive definite), the Maximum Likelihood estimator is consistent.

## (d) Estimation of $\alpha_i$

In a Bayesian (or Correlated Random Effects) framework, since we cannot estimate $\alpha_i$ consistently (it does not converge to a point), we estimate its **conditional posterior distribution** or its **conditional expectation (BLUP)** given the data.
The posterior kernel is:

$$\begin{aligned}
p(\alpha_i \mid \tilde{\mathbf{y}}_i, y_{i0}, \phi, \rho) &\propto p(\tilde{\mathbf{y}}_i \mid \alpha_i)\, p(\alpha_i \mid y_{i0}, \phi) \\
&\propto \exp\Big(-\tfrac{1}{2}\sum_{t=1}^{T}(\tilde{y}_{it} - \alpha_i)^2\Big) \cdot \exp\Big(-\tfrac{1}{2}(\alpha_i - \phi y_{i0})^2\Big) \\
&= \exp\Big(-\tfrac{1}{2}\big[c(\alpha_i - \tfrac{b}{c})^2 + a - \tfrac{b^2}{c}\big]\Big) \quad \text{(see definitions of } c, b, a \text{ above)}.
\end{aligned}$$

Therefore:

$$\alpha_i \mid \tilde{\mathbf{y}}_i, y_{i0}, \phi, \rho \;\sim\; \mathcal{N}\Big(\frac{\sum_{t=1}^{T}\tilde{y}_{it} + \phi y_{i0}}{T+1}, \frac{1}{T+1}\Big)$$

The posterior mean (Bayes estimator) is:

$$\begin{aligned}
\hat{\alpha}_i &= \mathbb{E}[\alpha_i \mid \tilde{\mathbf{y}}_i, y_{i0}, \phi, \rho] \\
&= \frac{\sum_{t=1}^{T}\tilde{y}_{it} + \phi y_{i0}}{T+1} \\
&= \frac{\sum_{t=1}^{T}(y_{it} - \rho y_{i,t-1}) + \phi y_{i0}}{T+1}.
\end{aligned}$$

we would first estimate $(\hat{\phi}, \hat{\rho})$ from the marginal likelihood, then plug them into the above expression to get $\hat{\alpha}_i$.

# Problem 2: State-Space Model

## Model

$$
\begin{aligned}
y_t &= \lambda s_t + u_t \\
s_t &= \phi s_{t-1} + \varepsilon_t \\
u_t &\sim \mathcal{N}(0,1), \quad \varepsilon_t \sim \mathcal{N}(0,1), \quad u_t \perp \varepsilon_t
\end{aligned}
$$

### (a) Autocovariance Function for $y_t$

Assuming stationarity ($|\phi| < 1$), the variance of the state $s_t$ is $\text{Var}(s_t) = \frac{1}{1-\phi^2}$. The covariance of the state is $\gamma_k^s = E[s_t s_{t-k}] = \phi^k \frac{1}{1-\phi^2}$.
For $y_t$:
**Variance ($\gamma_0$):**

$$\gamma_0 = E[(\lambda s_t + u_t)^2] = \lambda^2 \text{Var}(s_t) + \text{Var}(u_t) = \frac{\lambda^2}{1-\phi^2} + 1$$

**Autocovariance ($\gamma_k, k \geq 1$):**

$$
\begin{aligned}
\gamma_k &= E[y_t y_{t-k}] = E[(\lambda s_t + u_t)(\lambda s_{t-k} + u_{t-k})] \\
&= \lambda^2 E[s_t s_{t-k}] = \lambda^2 \frac{\phi^k}{1-\phi^2}
\end{aligned}
$$

Since $u_t$ is independent of $s_{t-k}$, $u_{t-k}$, and $s_t$ (for $k \geq 1$).

$$s_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j} \Rightarrow \mathbb{E}[s_t] = 0,\ \mathbb{V}[s_t] = \sum_{j\geq 0} \phi^{2j} = \frac{1}{1-\phi^2},$$

$$\text{Cov}(s_t, s_{t-k}) = \sum_{j\geq 0}\sum_{\ell \geq 0} \phi^{j+\ell}\, \text{Cov}(\varepsilon_{t-j}, \varepsilon_{t-k-\ell}) = \sum_{j\geq 0} \phi^{j+k+j} = \frac{\phi^k}{1-\phi^2}.$$

$$y_t = \lambda s_t + u_t \ \perp \ s_t \Rightarrow \gamma_0 = \mathbb{V}[y_t] = \lambda^2 \mathbb{V}[s_t] + 1 = 1 + \frac{\lambda^2}{1-\phi^2},$$

$$\gamma_k = \text{Cov}(y_t, y_{t-k}) = \lambda^2 \text{Cov}(s_t, s_{t-k}) = \frac{\lambda^2 \phi^k}{1-\phi^2} \quad (k \geq 1).$$

### (b) Identification

We have two unknown parameters $(\lambda, \phi)$ and we observe the autocovariances of $y$.

1. From $\gamma_1 = \frac{\lambda^2 \phi}{1-\phi^2}$ and $\gamma_0 = \frac{\lambda^2}{1-\phi^2} + 1$, notice that $\gamma_0 - 1 = \frac{\lambda^2}{1-\phi^2}$.

2. Thus, $\frac{\gamma_1}{\gamma_0 - 1} = \phi$.

3. Once $\phi$ is identified, $\lambda^2 = (\gamma_0 - 1)(1 - \phi^2)$.

**Result:** The coefficients are identified (up to the sign of $\lambda$, as only $\lambda^2$ enters the second moments).

## (c) ARMA Representation

From the state equation: $(1 - \phi L)s_t = \varepsilon_t \implies s_t = \frac{\varepsilon_t}{1-\phi L}$. Substitute into measurement equation:

$$y_t = \lambda \frac{\varepsilon_t}{1 - \phi L} + u_t$$

Multiply by $(1 - \phi L)$:

$$\begin{aligned}(1 - \phi L)y_t &= \lambda\varepsilon_t + (1 - \phi L)u_t \\ y_t - \phi y_{t-1} &= \lambda\varepsilon_t + u_t - \phi u_{t-1}\end{aligned}$$

Let the RHS be $w_t$. Since $w_t$ is a sum of MA processes, it is an MA(1) process $w_t = \nu_t + \theta\nu_{t-1}$. The LHS is AR(1). Thus, $y_t$ follows an **ARMA(1,1)** process. Parameters $(\phi_{AR}, \theta_{MA}, \sigma_\nu^2)$ are functions of $(\lambda, \phi, 1, 1)$. $\phi_{AR} = \phi$.

$$(1 - \phi L)y_t = \lambda\varepsilon_t + u_t - \phi u_{t-1} =: w_t.$$

Calculate the second moments of $w_t$:

$$\gamma_0^w = \mathbb{V}[w_t] = \lambda^2 + \mathbb{V}[u_t] + \phi^2\mathbb{V}[u_{t-1}] = \lambda^2 + 1 + \phi^2,$$

$$\gamma_1^w = \mathrm{Cov}(w_t, w_{t-1}) = \mathrm{Cov}(-\phi u_{t-1}, u_{t-1}) = -\phi.$$

For an MA(1) process: $w_t = v_t + \theta v_{t-1}$, where $v_t \sim \mathcal{N}(0, \sigma_v^2)$:

$$\gamma_0^w = (1 + \theta^2)\sigma_v^2, \quad \gamma_1^w = \theta\sigma_v^2.$$

Matching the equations:

$$\theta\sigma_v^2 = -\phi, \quad (1 + \theta^2)\sigma_v^2 = \lambda^2 + 1 + \phi^2$$

$$\Rightarrow -\phi(1 + \theta^2) = \theta(\lambda^2 + 1 + \phi^2) \Rightarrow \boxed{\phi\theta^2 + (\lambda^2 + 1 + \phi^2)\theta + \phi = 0}.$$

$$\sigma_v^2 = -\frac{\phi}{\theta}, \quad \text{choosing the root where } |\theta| < 1.$$

## (i) Allowing Correlation: $\mathrm{Cov}(u_t, \varepsilon_t) = \rho$

**Decomposition Method** (Moving the correlation to the measure in one step):

$$u_t = \rho\,\varepsilon_t + \eta_t, \qquad \eta_t \sim \mathcal{N}(0,,1-\rho^2),\ \eta_t \perp \varepsilon_s\ \forall s.$$

Thus:

$$y_t = \lambda s_t + \rho\varepsilon_t + \eta_t.$$

### ACF (Term-by-term)

$$\mathrm{Cov}(s_t, u_t) = \mathrm{Cov}\Big(\sum_{j\geq 0}\phi^j\varepsilon_{t-j},\ \rho\varepsilon_t + \eta_t\Big) = \rho,$$

$$\gamma_0 = \mathbb{V}[y_t] = \lambda^2\mathbb{V}[s_t] + \mathbb{V}[u_t] + 2\lambda\,\mathrm{Cov}(s_t,u_t) = \frac{\lambda^2}{1-\phi^2} + 1 + 2\lambda\rho,$$

$$\gamma_1 = \mathrm{Cov}(y_t, y_{t-1}) = \lambda^2\frac{\phi}{1-\phi^2} + \lambda\,\phi\rho, \qquad \gamma_k = \frac{\lambda^2\phi^k}{1-\phi^2}\ (k\geq 2).$$

### Equivalent ARMA(1,1) (Explicitly writing $w_t$)

$$(1-\phi L)y_t = (\lambda+\rho)\varepsilon_t + \eta_t - \phi\rho\,\varepsilon_{t-1} - \phi\,\eta_{t-1} =: w_t.$$

Second moments:

$$\gamma_0^w = \underbrace{(\lambda+\rho)^2 + \phi^2\rho^2}_{\text{from } \varepsilon} + \underbrace{(1-\rho^2)(1+\phi^2)}_{\text{from } \eta} = \lambda^2 + 1 + \phi^2 + 2\lambda\rho,$$

$$\gamma_1^w = \mathrm{Cov}(-\phi\rho\,\varepsilon_{t-1} - \phi\eta_{t-1},\ (\lambda+\rho)\varepsilon_{t-1} + \eta_{t-1}) = -\phi\big(\rho(\lambda+\rho) + (1-\rho^2)\big) = -\phi(1+\lambda\rho).$$

Matching with MA(1) moments:

$$\theta\sigma_v^2 = \gamma_1^w = -\phi(1+\lambda\rho), \quad (1+\theta^2)\sigma_v^2 = \gamma_0^w = \lambda^2 + 1 + \phi^2 + 2\lambda\rho,$$

$$\Rightarrow\ -\phi(1+\lambda\rho)(1+\theta^2) = \theta(\lambda^2 + 1 + \phi^2 + 2\lambda\rho),$$

$$\Rightarrow\ \boxed{\phi(1+\lambda\rho)\theta^2 + (\lambda^2 + 1 + \phi^2 + 2\lambda\rho)\theta + \phi(1+\lambda\rho) = 0},$$

$$\sigma_v^2 = -\frac{\phi(1+\lambda\rho)}{\theta}, \quad |\theta| < 1.$$

## (d) - (h) Code Implementation

Below is the Python code for the Kalman Filter, plotting, and optimization. Following that are the R and Julia translations.

```
set.seed(2025)
```

```r

simulate_ssm <- function(T=100, lam=1.0, phi=0.8, rho=0.0, seed=NULL){
  if(!is.null(seed)) set.seed(seed)

  Sigma <- matrix(c(1, rho, rho, 1), 2, 2)
  eps_u <- MASS::mvrnorm(T, mu=c(0,0), Sigma=Sigma)
  eps <- eps_u[,1]
  u <- eps_u[,2]

  s <- numeric(T)
  s[1] <- rnorm(1, 0, sqrt(1/(1-phi^2)))

  for(t in 2:T){
    s[t] <- phi * s[t-1] + eps[t]
  }

  y <- lam * s + u
  list(s=s, y=y, eps=eps, u=u)
}

kalman_filter <- function(y, lam, phi, rho=0.0){
  Tn <- length(y)

  m_pred <- P_pred <- m_filt <- P_filt <- loglik_t <- numeric(Tn)

  m_prev <- 0.0
  P_prev <- 1/(1-phi^2)

  for(t in 1:Tn){
    m_t_pred <- phi * m_prev
    P_t_pred <- phi^2 * P_prev + 1.0
    v_t <- y[t] - lam * m_t_pred
    F_t <- lam^2 * P_t_pred + 1.0 + 2 * lam * rho

    K_t <- (lam * P_t_pred + rho) / F_t

    m_t_filt <- m_t_pred + K_t * v_t
    P_t_filt <- P_t_pred - K_t * (lam * P_t_pred + rho)

    m_pred[t] <- m_t_pred    # E[s_t|Y_{1:t-1}]
    P_pred[t] <- P_t_pred    # V[s_t|Y_{1:t-1}]
    m_filt[t] <- m_t_filt    # E[s_t|Y_{1:t}]
    P_filt[t] <- P_t_filt    # V[s_t|Y_{1:t}]

    loglik_t[t] <- -0.5*(log(2*pi) + log(F_t) + v_t^2/F_t)

    m_prev <- m_t_filt
    P_prev <- P_t_filt
```

```r
  }

  list(m_pred=m_pred, P_pred=P_pred,
       m_filt=m_filt, P_filt=P_filt,
       loglik_t=loglik_t)
}

loglik_ssm <- function(y, lam, phi, rho=0.0){
  sum(kalman_filter(y, lam, phi, rho)$loglik_t)
}

# ========================
# (d)-(e)
# ========================

lam_true <- 0.9
phi_true <- 0.7
T_val <- 100

sim <- simulate_ssm(T=T_val, lam=lam_true, phi=phi_true, rho=0.0, seed
 =42)
y <- sim$y
s <- sim$s

out_true <- kalman_filter(y, lam=lam_true, phi=phi_true, rho=0.0)

out_wrong <- kalman_filter(y, lam=0.5, phi=0.3, rho=0.0)

tgrid <- 1:T_val
se_pred <- sqrt(out_true$P_pred)

plot(tgrid, s, type="l", lwd=2, col=1,
     ylab="State", xlab="Time",
     main="True vs Filtered States",
     ylim=range(s, out_true$m_pred, out_wrong$m_pred))
lines(tgrid, out_true$m_pred, col="blue", lwd=1.5)
lines(tgrid, out_true$m_pred + 1.96*se_pred, col="blue", lty=2, lwd=1)
lines(tgrid, out_true$m_pred - 1.96*se_pred, col="blue", lty=2, lwd=1)
lines(tgrid, out_wrong$m_pred, col="red", lwd=1.5, lty=3)
legend("topleft",
       c("True state s_t",
         "Filtered E[s_t|Y_{1:t-1}] (true params)",
         "95% prediction bands",
         "Filtered (wrong params)"),
       col=c(1, "blue", "blue", "red"),
       lty=c(1,1,2,3), lwd=c(2,1.5,1,1.5),
       bty="n", cex=0.8)

```

```r
plot(tgrid, exp(out_true$loglik_t), type="l", lwd=2, col="blue",
     ylab="Likelihood increment p(y_t|Y_{1:t-1})",
     xlab="Time", main="Likelihood Increments")
lines(tgrid, exp(out_wrong$loglik_t), col="red", lwd=1.5, lty=2)
legend("topright", c("True parameters", "Wrong parameters"),
        col=c("blue", "red"), lty=1:2, lwd=2:1.5, bty="n")
```

```r

# ========================
# (f)
# ========================

phi_grid <- seq(0.01, 0.99, length.out=200)
ll_grid <- sapply(phi_grid, function(p) loglik_ssm(y, lam=lam_true,
  phi=p))

phi_hat_grid <- phi_grid[which.max(ll_grid)]
cat(sprintf("Grid search MLE: \phi_hat = %.4f (true \phi = %.2f)\n",
phi_hat_grid, phi_true))

plot(phi_grid, ll_grid, type="l", lwd=2,
xlab="\phi", ylab="Log-likelihood",
main="Profile Log-likelihood for \phi")
abline(v=phi_true, col="green", lty=2, lwd=1.5)
abline(v=phi_hat_grid, col="red", lty=3, lwd=1.5)
legend("topright", c("Log-likelihood", "True \phi", "MLE \phi"),
col=c(1, "green", "red"), lty=c(1,2,3), lwd=c(2,1.5,1.5), bty="n")
```

```r

# ========================
# (g)
# ========================

sample_sizes <- c(50, 100, 500)
colors <- c("red", "blue", "green")

plot(phi_grid, ll_grid, type="n",
     xlab="\phi", ylab="Log-likelihood",
     main="Log-likelihood for Different Sample Sizes",
     ylim=c(-1000, 0))

for (i in 1:length(sample_sizes)) {
  T_i <- sample_sizes[i]
  sim_i <- simulate_ssm(T=T_i, lam=lam_true, phi=phi_true, seed=100+T_i)
  ll_i <- sapply(phi_grid, function(p) loglik_ssm(sim_i$y, lam=lam_true,
    phi=p))
  lines(phi_grid, ll_i, col=colors[i], lwd=1.5)
}

```

```
legend("bottomleft", legend=paste0("T=", sample_sizes),
       col=colors, lty=1, lwd=1.5, bty="n")
```

```
  # ==========================
  # (h)
  # ==========================

  neg_loglik <- function(phi) {
      if(phi <= 0 || phi >= 1) return(Inf)
      -loglik_ssm(y, lam=lam_true, phi=phi)
    }

  opt_result <- optimize(neg_loglik, interval=c(0.01, 0.99))
  phi_hat_opt <- opt_result$minimum

  cat(sprintf("Numerical optimization:\n"))
  cat(sprintf("  \phi_hat (optim) = %.6f\n", phi_hat_opt))
  cat(sprintf("  \phi_hat (grid)  = %.6f\n", phi_hat_grid))
  cat(sprintf("  Difference     = %.6f\n", abs(phi_hat_opt - phi_hat_grid)))
  cat(sprintf("  Log-lik at opt = %.4f\n", -opt_result$objective))
```

```

# ==========================
# (i)-(j)
# ==========================

rho_val <- 0.5
sim_corr <- simulate_ssm(T=100, lam=lam_true, phi=phi_true, rho=rho_val, seed=123)

acf_y <- acf(sim_corr$y, lag.max=10, plot=FALSE)
cat("\nSample autocovariances (with correlation):\n")
for(k in 0:3) {
  cat(sprintf("\gamma [%d] = %.4f\n", k, acf_y$acf[k+1]*var(sim_corr$y)))
}

kf_corr <- kalman_filter(sim_corr$y, lam=lam_true, phi=phi_true, rho=rho_val)

kf_wrong <- kalman_filter(sim_corr$y, lam=lam_true, phi=phi_true, rho=0.0)

plot(tgrid, sim_corr$s, type="l", lwd=2, col=1,
     ylab="State", xlab="Time",
     main="KF with Correlated Errors",
     ylim=range(sim_corr$s, kf_corr$m_pred, kf_wrong$m_pred))
lines(tgrid, kf_corr$m_pred, col="blue", lwd=1.5)
```

```r
lines(tgrid, kf_wrong$m_pred, col="red", lwd=1.5, lty=2)
legend("topleft",
       c("True state", "KF with correct rho", "KF assuming rho=0"),
       col=c(1, "blue", "red"), lty=c(1,1,2), lwd=c(2,1.5,1.5),
       bty="n", cex=0.8)

ll_corr <- sum(kf_corr$loglik_t)
ll_wrong <- sum(kf_wrong$loglik_t)
cat(sprintf("\nLog-likelihood comparison:\n"))
cat(sprintf("  With correct rho=%.2f: %.4f\n", rho_val, ll_corr))
cat(sprintf("  Assuming rho=0:        %.4f\n", ll_wrong))
cat(sprintf("  Difference:            %.4f\n", ll_corr - ll_wrong))

par(mfrow=c(1,1))
```

## (i) Correlated Errors

Suppose $\text{Cov}(u_t, \varepsilon_t) = \rho$.
**Autocovariance:** $\gamma_0 = \lambda^2 Var(s_t) + Var(u_t) + 2\lambda Cov(s_t, u_t)$. Since $s_t = \phi s_{t-1} + \varepsilon_t$, $Cov(s_t, u_t) = Cov(\varepsilon_t, u_t) = \rho$.

$$\gamma_0 = \frac{\lambda^2}{1-\phi^2} + 1 + 2\lambda\rho$$

$\gamma_1 = E[(\lambda s_t + u_t)(\lambda s_{t-1} + u_{t-1})]$. $E[u_t s_{t-1}] = 0$, $E[u_t u_{t-1}] = 0$. $E[s_t u_{t-1}] = E[(\phi s_{t-1} + \varepsilon_t)u_{t-1}] = \phi Cov(s_{t-1}, u_{t-1}) = \phi\rho$.

$$\gamma_1 = \lambda^2 \phi Var(s_t) + \lambda E[s_t u_{t-1}] = \lambda^2 \frac{\phi}{1-\phi^2} + \lambda\phi\rho$$

**Identification:** Yes, if moments differ, though the mapping is more complex.
**ARMA:** Still ARMA(1,1) because it is the sum of two correlated processes, one AR(1) and one White Noise. The spectral density will maintain the rational form.

## (j) Generalized Kalman Filter with Correlation

If $E[u_t\varepsilon_t] = \rho \neq 0$, the innovation in the measurement ($u_t$) contains information about the innovation in the state ($\varepsilon_t$). Standard KF Prediction step ($s_{t|t-1} \to s_{t+1|t}$) must change. The posterior of the state $s_t$ given $y_t$ updates as usual, but when predicting $s_{t+1} = \phi s_t + \varepsilon_t$, we must note that $\varepsilon_t$ is correlated with the measurement error $u_t$ contained in $y_t$.
**Modified Algorithm:**

1. **State Prediction:** $s_{t|t-1}$ (Same)
2. **Measurement Prediction:** $y_{t|t-1} = \lambda s_{t|t-1}$. Error $v_t = y_t - y_{t|t-1}$.

3. **Covariance of Innovation:**

$$Cov(s_{t+1}, y_t|t-1) = E[(\phi(s_t - s_{t|t-1}) + \varepsilon_t)(\lambda(s_t - s_{t|t-1}) + u_t)] = \phi\lambda P_{t|t-1} + \rho$$

   *(Note the addition of $\rho$).*

4. **Kalman Gain:**

$$K_t = (\phi\lambda P_{t|t-1} + \rho)F_t^{-1}$$

5. **State Update (Predict next step directly):**

$$\begin{aligned} s_{t+1|t} &= \phi s_{t|t-1} + K_t v_t \\ P_{t+1|t} &= \phi^2 P_{t|t-1} + 1 - K_t F_t K_t' \end{aligned}$$

   *(The standard KF separates update $t|t$ and predict $t+1|t$, but with correlation it is often cleaner to write the one-step ahead recursion directly).*

Given $Y_{1:t-1}$, denote:

$$m_{t|t-1} = \mathbb{E}[s_t \mid Y_{1:t-1}], \quad P_{t|t-1} = \mathbb{V}[s_t \mid Y_{1:t-1}].$$

State and Observation:

$$s_t = \phi s_{t-1} + \varepsilon_t, \quad y_t = \lambda s_t + u_t, \quad \mathrm{Cov}(\varepsilon_t, u_t) = \rho.$$

**1) Prediction:**

$$m_{t|t-1} = \phi\, m_{t-1|t-1}, \qquad P_{t|t-1} = \phi^2 P_{t-1|t-1} + 1.$$

**2) Innovation and its Variance (Explicit term-wise covariance):**

$$v_t \equiv y_t - \lambda m_{t|t-1} = \lambda(s_t - m_{t|t-1}) + u_t.$$

Thus:

$$\begin{aligned} F_t &= \mathbb{V}[v_t \mid Y_{1:t-1}] = \mathbb{V}[\lambda(s_t - m_{t|t-1}) + u_t] \\ &= \lambda^2\, \mathbb{V}[s_t - m_{t|t-1}] + \mathbb{V}[u_t] + 2\lambda\, \mathrm{Cov}(s_t - m_{t|t-1}, u_t) \\ &= \lambda^2 P_{t|t-1} + 1 + 2\lambda\, \mathrm{Cov}(s_t, u_t). \end{aligned}$$

Also:

$$\mathrm{Cov}(s_t, u_t) = \mathrm{Cov}(\phi s_{t-1} + \varepsilon_t,\ u_t) = \mathrm{Cov}(\varepsilon_t, u_t) = \rho,$$

So:

$$\boxed{F_t = \lambda^2 P_{t|t-1} + 1 + 2\lambda\rho}.$$

**3) Gain (Using "Cross-Covariance / Innovation Variance"):**

$$\begin{aligned} K_t &= \frac{\operatorname{Cov}(s_t, v_t \mid Y_{1:t-1})}{F_t} = \frac{\operatorname{Cov}(s_t, \lambda(s_t - m_{t|t-1}) + u_t)}{F_t} \\ &= \frac{\lambda\,\mathbb{V}[s_t - m_{t|t-1}] + \operatorname{Cov}(s_t, u_t)}{F_t} = \boxed{\frac{\lambda P_{t|t-1} + \rho}{F_t}}. \end{aligned}$$

**4) Update:**

$$m_{t|t} = m_{t|t-1} + K_t v_t, \qquad P_{t|t} = P_{t|t-1} - K_t \operatorname{Cov}(s_t, v_t \mid Y_{1:t-1}) = P_{t|t-1} - K_t(\lambda P_{t|t-1} + \rho).$$

**5) Log-Likelihood Increment:**

$$\log p(y_t \mid Y_{1:t-1}) = -\tfrac{1}{2}\Big(\log 2\pi + \log F_t + \frac{v_t^2}{F_t}\Big).$$