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.

Manifestation: In a dynamic panel (where y_{it-1} is a regressor), the standard Fixed Effects (Within) estimator or the naive MLE for α_i and ρ yields inconsistent estimates for ρ when $N \to \infty$ but T remains fixed. This happens because the estimation error of α_i (which does not vanish as $N \to \infty$) is correlated with the regressor y_{it-1} (since y_{it-1} contains α_i). This induces a downward bias in the estimate of ρ, commonly known as the **Nickell Bias** (of order $1/T$).

(b) Integrating out α_i

To integrate out α_i, we view the model vector-wise for individual i. Let $\tilde{y}_{it} = y_{it} - \rho y_{it-1}$. The structural equation becomes:

$$\tilde{y}_{it} = \alpha_i + u_{it}$$

In vector notation for $t = 1 : T$:

$$\tilde{y}_i = \mathbf{1}_T \alpha_i + u_i$$

where $\mathbf{1}_T$ is a column vector of ones.

We are given $\alpha_i = \phi y_{i0} + \eta_i$, where $\eta_i \sim \mathcal{N}(0,1)$. Substituting this into the vector equation:

$$\tilde{y}_i = \mathbf{1}_T(\phi y_{i0} + \eta_i) + u_i$$
$$\tilde{y}_i = \mathbf{1}_T \phi y_{i0} + (\mathbf{1}_T \eta_i + u_i)$$

The composite error term is $v_i = \mathbf{1}_T \eta_i + u_i$. We compute the mean and variance of $\tilde{y}_i$ conditional on y_{i0}:

1. **Mean:** $E[\tilde{y}_i | y_{i0}] = \mathbf{1}_T \phi y_{i0}$
2. **Variance:** $\Omega = \text{Var}(v_i) = E[(\mathbf{1}_T\eta_i + u_i)(\mathbf{1}_T\eta_i + u_i)'] = \mathbf{1}_T \mathbf{1}_T' \text{Var}(\eta_i) + \text{Var}(u_i)$. Since $\text{Var}(\eta_i) = 1$ and $\text{Var}(u_i) = I_T$:

$$\Omega = \mathbf{1}_T \mathbf{1}_T' + I_T$$

The marginal likelihood function $p(y_{i1}, \ldots, y_{iT} | y_{i0}, \phi, \rho)$ is the multivariate normal density of $\tilde{y}_i$ evaluated at the observed data (transformed by ρ), with mean $\mathbf{1}_T \phi y_{i0}$ and covariance Ω:

$$p(y_i | y_{i0}, \phi, \rho) = (2\pi)^{-\frac{T}{2}} |\Omega|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (\tilde{y}_i - \mathbf{1}_T \phi y_{i0})' \Omega^{-1} (\tilde{y}_i - \mathbf{1}_T \phi y_{i0}) \right)$$

Note: $\tilde{y}_i$ depends on ρ.

(c) Consistency of (ϕ, ρ)

Yes, (ϕ, ρ) can be consistently estimated. By integrating out α_i, we have removed the incidental parameters. We are left with a likelihood function that depends on a finite number of common parameters (ϕ, ρ) and data vectors y_i. Assuming cross-sectional independence (observations $i = 1 : N$ are i.i.d.), the log-likelihood for the whole sample scales with N:

$$\mathcal{L}(\phi, \rho) = \sum_{i=1}^{N} \ln p(y_i | y_{i0}, \phi, \rho)$$

Standard Maximum Likelihood theory applies: as $N \to \infty$ (even with fixed T), the estimator maximizing this marginal likelihood is consistent and asymptotically normal, provided identification conditions hold.

(d) Estimation of α_i

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

Using Bayes' rule: $p(\alpha_i|y_i, y_{i0}) \propto p(y_i|\alpha_i, \dots)p(\alpha_i|y_{i0})$. Given the normal-normal conjugacy, the estimator would be the posterior mean:

$$E[\alpha_i|y_i, y_{i0}] = \hat{\alpha}_i = w\bar{\tilde{y}}_i + (1-w)\phi y_{i0}$$

where $\bar{\tilde{y}}_i$ is the mean of residuals $y_{it} - \rho y_{it-1}$ and w is a shrinkage factor depending on the relative precision of the signal T/σ_u^2 and the prior precision $1/\sigma_\alpha^2$.

Problem 2: State-Space Model

Model

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

(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 (γ_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 $k \geq 1$).

(b) Identification

We have two unknown parameters (λ, ϕ) 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 ϕ is identified, $\lambda^2 = (\gamma_0 - 1)(1-\phi^2)$.

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

(c) ARMA Representation

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

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

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

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

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 $(\lambda, \phi, 1, 1)$. $\phi_{AR} = \phi$.

(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)

# Simulate Data
simulate_data <- function(T, lam, phi) {
  s <- numeric(T)
  y <- numeric(T)
  s[1] <- rnorm(1, 0, sqrt(1/(1-phi^2)))
  y[1] <- lam * s[1] + rnorm(1)

  eps <- rnorm(T)
  u <- rnorm(T)

  for(t in 2:T) {
    s[t] <- phi * s[t-1] + eps[t]
    y[t] <- lam * s[t] + u[t]
  }
  return(list(y=y, s=s))
}

# Kalman Filter
kalman_filter <- function(Y, lam, phi) {
  T <- length(Y)
  s_pred <- numeric(T)
```

```
  P_pred <- numeric(T)
  s_upd <- numeric(T)
  P_upd <- numeric(T)
  ll_contrib <- numeric(T)

  # Initialization
  s_pred[1] <- 0
  P_pred[1] <- 1 / (1 - phi^2)

  for(t in 1:T) {
    # Prediction error decomp
    y_pred <- lam * s_pred[t]
    F_t <- lam^2 * P_pred[t] + 1
    v_t <- Y[t] - y_pred

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

    # Update
    K_t <- P_pred[t] * lam / F_t
    s_upd[t] <- s_pred[t] + K_t * v_t
    P_upd[t] <- P_pred[t] * (1 - K_t * lam)

    # Predict next
    if(t < T) {
      s_pred[t+1] <- phi * s_upd[t]
      P_pred[t+1] <- phi^2 * P_upd[t] + 1
    }
  }
  return(list(ll=ll_contrib, s_pred=s_pred, P_pred=P_pred))
}

# Analysis
T <- 100
true_lam <- 1.0
true_phi <- 0.8
data <- simulate_data(T, true_lam, true_phi)

kf_res <- kalman_filter(data$y, true_lam, true_phi)

# Optimization
neg_ll <- function(p) {
  if(abs(p) >= 0.999) return(Inf)
  -sum(kalman_filter(data$y, true_lam, p)$ll)
}

opt_res <- optim(0.5, neg_ll, method="L-BFGS-B", lower=-0.99, upper
   =0.99)
```

```
print(paste("Optimization Estimate:", opt_res$par))
```

(i) Correlated Errors

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

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

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

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

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 $E[u_t\epsilon_t] = \rho \neq 0$, the innovation in the measurement (u_t) contains information about the innovation in the state (ϵ_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 $s_{t+1} = \phi s_t + \epsilon_t$, we must note that ϵ_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:**

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

(Note the addition of ρ).

4. **Kalman Gain:**

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

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

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

(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).