Model Estimation with AR(1) Endogenous Unobservable

1) Model:

The observables are y_{it} and x_{it}. The goal is to identify β in the relationship,

$$y_{it} = x_{it}'\beta + w_{it}'\gamma + \epsilon_{it} + \nu_{it}, \quad \nu_{it} \sim \text{iid}, \quad \nu_{it} \text{ is serially independent} \tag{1}$$

$$x_{it} = \gamma\epsilon_{it} + (1-\gamma)\, u_{it}, \quad u_{it} \sim \text{iid}, \quad u_{it} \text{ is serially dependent} \tag{2}$$

$$\epsilon_{it} = \rho\epsilon_{it-1} + \eta_{it}, \quad \eta_{it} \sim \text{iid}, \quad \eta_{it} \text{ is serially independent} \tag{3}$$

where w_{it} includes a constant.

2) Identification:

The usual argument for identification is based on the quasi-difference expression:

$$(y_{it} - \rho y_{it-1}) = (x_{it} - \rho x_{it-1})'\,\beta + (w_{it} - \rho w_{it-1})'\,\delta + \eta_{it} \tag{4}$$

GMM Approach: For any guess $\left(\hat{\beta}, \hat{\delta}, \hat{\rho}\right)$, we can define the guess of η_{it}:

$$\hat{\eta}_{it}\left(\hat{\beta}, \hat{\delta}, \hat{\rho}\right) \equiv (y_{it} - \hat{\rho} y_{it-1}) - (x_{it} - \hat{\rho} x_{it-1})'\,\hat{\beta} - (w_{it} - \rho w_{it-1})'\,\hat{\delta} \tag{5}$$

Then,

$$(\beta, \delta, \rho) \quad \text{solves} \quad \mathbb{E}\left[(w_{it}, z_{it})\ \hat{\eta}_{it}\left(\hat{\beta}, \hat{\delta}, \hat{\rho}\right)\right] = 0 \tag{6}$$

where z_{it} includes variables that are independent of η_{it}. The standard choices of instruments are $z_{it} = (x_{it-1}, y_{it-1})$ and $z_{it} = (x_{it-1}, x_{it-2})$. Note that the number of instruments z_{it} needs to be one more than the number of enodgenous variables in x_{it}.

Panel IV Approach: An alternative approach is to rearrange Equation (4) as a panel regression:

$$y_{it} = y_{it-1}(\rho) + x_{it}'(\beta) + x_{it-1}'(-\rho\beta) + w_{it}'(\delta) + w_{it-1}'(-\rho\delta) + \eta_{it} \tag{7}$$

The only source of endogeneity in this regression is that x_{it} depends on η_{it}. This implies that β is identified by a regression of y_{it} on x_{it}, controlling for $(y_{it-1}, x_{it-1}, w_{it}, w_{it-1})$, and instrumented by $z_{it} = (x_{it-2})$ or $z_{it} = (x_{it-2}, y_{it-2})$.

3) Simulation Exercise:

In order to compare the estimation approaches, Figure 1 simulates the model defined above. It sets $\beta = (0.5, -0.2)$, $\delta = 1$, $\rho = 0.5$, $\gamma = 0.5$, $\eta_{it} \sim \mathcal{N}(0,1)$, $u_{it} = u_{it-1} + \mathcal{N}(0,1)$, and $\nu_{it} = 0$. The length of the panel is $T = 3$, which is the minimum required. For various choices of N, it draws 10 random samples from the model, applies the estimators, and presents the box-plot of the distribution of estimates.

4) Controlling for Covariates

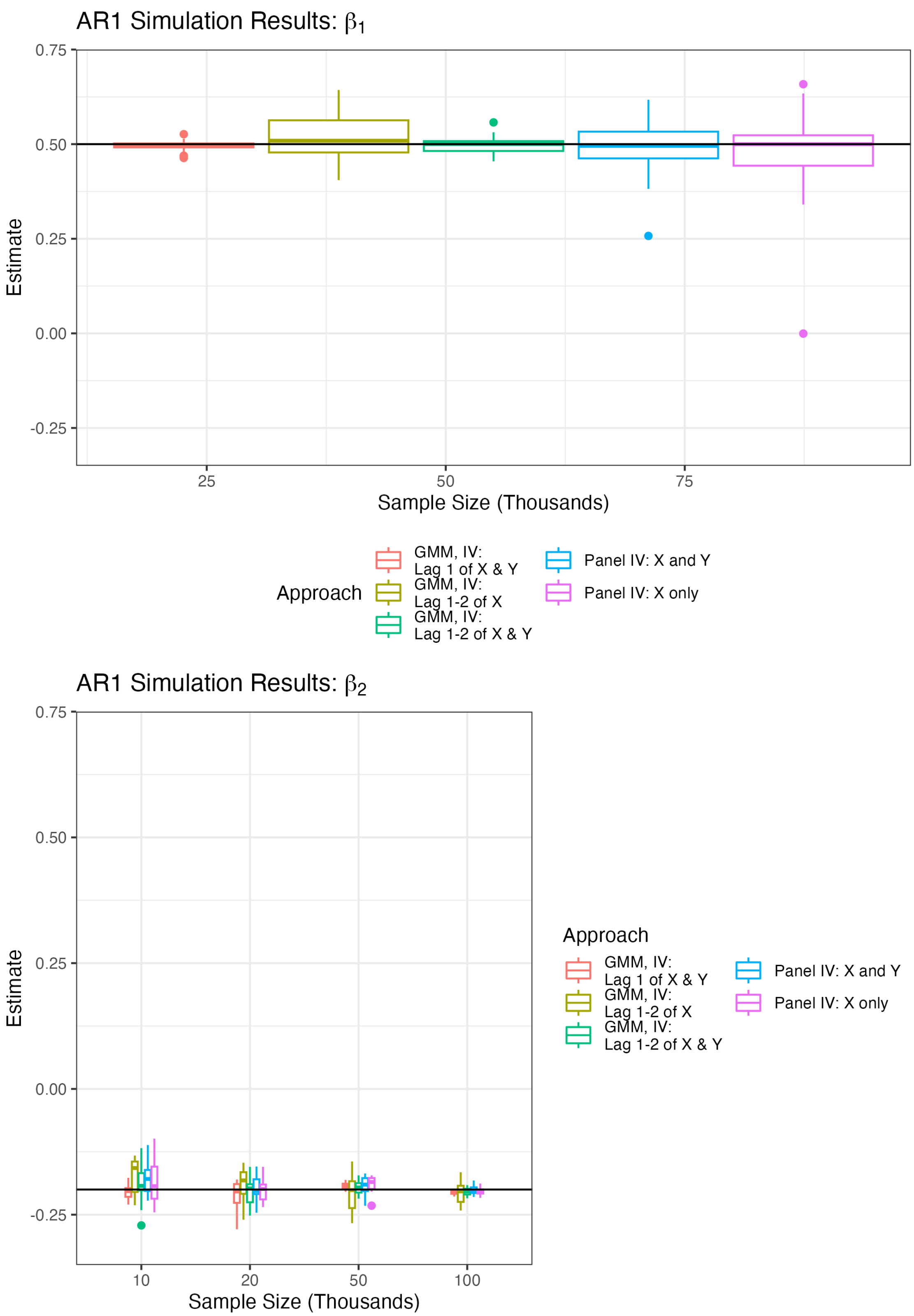

Figure 1: Simulation Exercise