PS4 - Proposed Solutions

Taric Latif Padovani

November 10, 2025

Problem 1. 1. We are given a strictly (weakly) stationary $\{X_t\}$ with $\mathbb{E}[X_t] = 0$, covariance function $\gamma(h) = (X_t, X_{t-h})$, and the process is α–mixing. Define

$$Y_t \; = \; X_t + 0.5\, X_{t-1}, \qquad t \in \mathbb{Z}.$$

Because Y_t is a finite linear combination of (X_t), it is also weakly stationary and (under the standard fact that finite linear transforms preserve mixing with the same rate) α–mixing.

To apply a CLT for strongly/alpha–mixing sequences (e.g. Ibragimov–Linnik), we *impose* the usual additional conditions:

- there exists $\delta > 0$ such that $\mathbb{E}|X_0|^{2+\delta} < \infty$ (hence also $\mathbb{E}|Y_0|^{2+\delta} < \infty$);
- the mixing coefficients satisfy $\sum_{h=1}^{\infty} \alpha(h)^{\delta/(2+\delta)} < \infty$.

Under these assumptions, the CLT for stationary α–mixing sequences gives

$$\sqrt{T}\Big(\bar{Y}_T - \mathbb{E}Y_0\Big) \xrightarrow{d} \mathcal{N}(0, \Omega_Y), \qquad \bar{Y}_T := \frac{1}{T}\sum_{t=2}^{T} Y_t,$$

and since $\mathbb{E}Y_0 = \mathbb{E}X_0 + 0.5\, \mathbb{E}X_{-1} = 0$, this is

$$\sqrt{T}\, \frac{1}{T}\sum_{t=2}^{T} Y_t \xrightarrow{d} \mathcal{N}(0, \Omega_Y).$$

(The fact that the sum starts at $t = 2$ instead of $t = 1$ is $o_p(1)$.)

It remains to identify the long–run variance. For $b_0 = 1$ and $b_1 = 0.5$ we can write

$$Y_t = \sum_{j=0}^{1} b_j X_{t-j}.$$

Hence, for $h \in \mathbb{Z}$,

$$\gamma_Y(h) := (Y_t, Y_{t-h}) = \sum_{j=0}^{1}\sum_{k=0}^{1} b_j b_k\, \gamma(h + j - k).$$

Plugging $b_0 = 1$, $b_1 = 0.5$,

$$\gamma_Y(h) = 1{\cdot}1{\cdot}\gamma(h) + 1{\cdot}0.5{\cdot}\gamma(h+1) + 0.5{\cdot}1{\cdot}\gamma(h-1) + 0.5{\cdot}0.5{\cdot}\gamma(h) = 1.25\,\gamma(h) + 0.5\,\gamma(h+1) + 0.5\,\gamma(h-1).$$

The long–run variance is then

$$\Omega_Y \;=\; \sum_{h=-\infty}^{\infty} \gamma_Y(h) = \gamma_Y(0) + 2\sum_{h=1}^{\infty} \gamma_Y(h),$$

with

$$\gamma_Y(0) = \mathrm{Var}(Y_t) = 1.25\,\gamma(0) + \gamma(1).$$

Therefore, under the mixing and moment assumptions we imposed,

$$\boxed{\sqrt{T}\,\frac{1}{T}\sum_{t=2}^{T} Y_t \xrightarrow{d} \mathcal{N}\big(0,\ \Omega_Y\big), \quad \Omega_Y = \gamma_Y(0) + 2\sum_{h=1}^{\infty}\big[1.25\,\gamma(h) + 0.5\,\gamma(h+1) + 0.5\,\gamma(h-1)\big].}$$

2. Now X_t is i.i.d., scalar, with $\mathbb{E}X_t = 0$ and $\mathbb{E}X_t^4 < \infty$. Define

$$Y_t := X_t X_{t-11}.$$

Because the X_t's are i.i.d., (Y_t) is a *finite–dependence* (in fact, 11–dependence) process: for $h = 1, \dots, 10$ the sets of indices $\{t, t-11\}$ and $\{t-h, t-h-11\}$ are disjoint, so Y_t and Y_{t-h} are independent. Only at lag $h = 11$ is there overlap, but even there the covariance will turn out to be zero. A classical CLT for m–dependent sequences (e.g. Billingsley, or the standard m–dependent CLT) applies if $\mathbb{E}Y_t^2 < \infty$, which we *do* have because

$$\mathbb{E}Y_t^2 = \mathbb{E}(X_t^2 X_{t-11}^2) = \mathbb{E}(X_t^2)\,\mathbb{E}(X_{t-11}^2) = \big(\mathbb{E}X_t^2\big)^2 < \infty$$

(using independence and the given finite fourth moment).

First, the mean:

$$\mathbb{E}Y_t = \mathbb{E}(X_t)\,\mathbb{E}(X_{t-11}) = 0.$$

Second, the covariances. For $h = 1, \dots, 10$ we have independence $\Rightarrow$ $(Y_t, Y_{t-h}) = 0$. For $h = 11$,

$$Y_t = X_t X_{t-11}, \qquad Y_{t-11} = X_{t-11} X_{t-22},$$

so

$$(Y_t, Y_{t-11}) = \mathbb{E}\big[X_t X_{t-11}^2 X_{t-22}\big] - \mathbb{E}Y_t \mathbb{E}Y_{t-11} = \mathbb{E}X_t\,\mathbb{E}X_{t-11}^2\,\mathbb{E}X_{t-22} = 0,$$

because $\mathbb{E}X_t = 0$. Thus *all* off–diagonal covariances are zero, and the long–run variance is just

$$\Omega = \mathrm{Var}(Y_t) = \mathbb{E}(X_t^2 X_{t-11}^2) = \big(\mathbb{E}X_t^2\big)^2.$$

We sum from $t = 2$ to T in the statement, but for this process the first 11 terms are "incomplete". Replacing $\frac{1}{T}\sum_{t=2}^{T}$ with $\frac{1}{T}\sum_{t=12}^{T}$ changes the average by $O(T^{-1})$, which is negligible at $\sqrt{T}$–rate. Hence, by the m–dependent CLT,

$$\sqrt{T}\,\frac{1}{T}\sum_{t=2}^{T} Y_t \xrightarrow{d} \mathcal{N}\big(0,\ (\mathbb{E}X_t^2)^2\big).$$

So the CLT we can state is

$$\boxed{\sqrt{T}\,\frac{1}{T}\sum_{t=2}^{T} X_t X_{t-11} \;\xrightarrow{d}\; \mathcal{N}\big(0,\ (\mathbb{E}X_t^2)^2\big),}$$

where we used the fact that (Y_t) is 11–dependent with zero mean and finite second moment, and we imposed no extra mixing assumptions beyond the i.i.d. structure.

Problem 2. We are given

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \end{bmatrix}, \qquad \mathbf{y}_{t-1} = \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{bmatrix}.$$

---

1. Show that $\mathbf{B}\mathbf{y}_{t-1}$ has the desired form.
Compute the product row by row.
- First row of $\mathbf{B}$ is $(1, 0, \ldots, 0)$, so

$$(\text{row } 1) \cdot \mathbf{y}_{t-1} = 1 \cdot y_{t-1} = y_{t-1}.$$

- Second row is $(1, -1, 0, \ldots, 0)$, so

$$(\text{row } 2) \cdot \mathbf{y}_{t-1} = y_{t-1} - y_{t-2} = \Delta y_{t-1}.$$

- Third row is $(0, 1, -1, 0, \ldots, 0)$, so

$$(\text{row } 3) \cdot \mathbf{y}_{t-1} = y_{t-2} - y_{t-3} = \Delta y_{t-2}.$$

- Continuing this pattern, the k-th nonzero row (for $k = 2, \ldots, p$) is zero everywhere except a 1 and a -1 in consecutive positions, giving

$$y_{t-(k-1)} - y_{t-k} = \Delta y_{t-k+1}.$$

Thus

$$\mathbf{B}\mathbf{y}_{t-1} = \begin{bmatrix} y_{t-1} \\ \Delta y_{t-1} \\ \Delta y_{t-2} \\ \vdots \\ \Delta y_{t-p+1} \end{bmatrix},$$

as required.

Now let

$$\phi = (\phi_1, \phi_2, \ldots, \phi_p)' \in \mathbb{R}^p, \qquad \begin{bmatrix} \rho \\ \beta \end{bmatrix} = \begin{bmatrix} \rho \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \in \mathbb{R}^p,$$

and assume

$$\phi = \mathbf{B}' \begin{bmatrix} \rho \\ \beta \end{bmatrix}.$$

We need to work out the explicit form of $\mathbf{B}'$.

Since

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \end{bmatrix},$$

its transpose is

$$\mathbf{B}' = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ 0 & 0 & -1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 0 & 0 & 0 & \cdots & -1 \end{bmatrix}.$$

So

$$\phi = \mathbf{B}' \begin{bmatrix} \rho \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{p-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ 0 & 0 & -1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 0 & 0 & 0 & \cdots & -1 \end{bmatrix} \begin{bmatrix} \rho \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{p-1} \end{bmatrix}.$$

Carrying out the multiplication componentwise:
- First component:

$$\phi_1 = \rho + \beta_1.$$

- Second component (use row $(0, -1, 1, 0, \dots)$):

$$\phi_2 = -\beta_1 + \beta_2.$$

- Third component:

$$\phi_3 = -\beta_2 + \beta_3.$$

- Continue this pattern up to

$$\phi_{p-1} = -\beta_{p-2} + \beta_{p-1},$$

and the last row of $\mathbf{B}'$ is $(0, 0, \dots, 0, -1)$, so

$$\phi_p = -\beta_{p-1}.$$

So we have the system

$$\begin{aligned} \phi_1 &= \rho + \beta_1, \\ \phi_2 &= -\beta_1 + \beta_2, \\ \phi_3 &= -\beta_2 + \beta_3, \\ &\vdots \\ \phi_{p-1} &= -\beta_{p-2} + \beta_{p-1}, \\ \phi_p &= -\beta_{p-1}. \end{aligned}$$

(a) Show that $\rho = \sum_{j=1}^{p} \phi_j$.
Sum all p equations above:

$$\sum_{j=1}^{p} \phi_j = (\rho + \beta_1) + (-\beta_1 + \beta_2) + (-\beta_2 + \beta_3) + \cdots + (-\beta_{p-2} + \beta_{p-1}) + (-\beta_{p-1}).$$

This is a telescoping sum: every $+\beta_k$ is canceled by a $-\beta_k$ in the next term, and the final $-\beta_{p-1}$ cancels the last $+\beta_{p-1}$. What remains is

$$\sum_{j=1}^{p} \phi_j = \rho.$$

Hence

$$\boxed{\rho = \sum_{j=1}^{p} \phi_j.}$$

(b) Rewrite the AR(p) in ADF form.
Start from the usual AR(p):

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t = \phi' \mathbf{y}_{t-1} + \varepsilon_t.$$

By assumption, $\phi = \mathbf{B}' \begin{bmatrix} \rho \\ \beta \end{bmatrix}$, so

$$\phi' \mathbf{y}_{t-1} = \left(\mathbf{B}'[\rho;\beta]\right)' \mathbf{y}_{t-1} = [\rho;\beta]' \mathbf{B} \mathbf{y}_{t-1}.$$

But from the first part we already proved

$$\mathbf{B}\mathbf{y}_{t-1} = \begin{bmatrix} y_{t-1} \\ \Delta y_{t-1} \\ \Delta y_{t-2} \\ \vdots \\ \Delta y_{t-p+1} \end{bmatrix}.$$

Therefore,

$$\phi' \mathbf{y}_{t-1} = [\rho; \beta_1; \ldots; \beta_{p-1}]' \begin{bmatrix} y_{t-1} \\ \Delta y_{t-1} \\ \Delta y_{t-2} \\ \vdots \\ \Delta y_{t-p+1} \end{bmatrix} = \rho y_{t-1} + \beta_1 \Delta y_{t-1} + \beta_2 \Delta y_{t-2} + \cdots + \beta_{p-1} \Delta y_{t-p+1}.$$

Substitute back into the AR(p) equation:

$$y_t = \rho y_{t-1} + \beta_1 \Delta y_{t-1} + \beta_2 \Delta y_{t-2} + \cdots + \beta_{p-1} \Delta y_{t-p+1} + \varepsilon_t.$$

So we have shown

$$\boxed{y_t = \rho y_{t-1} + \sum_{j=1}^{p-1} \beta_j \Delta y_{t-j} + \varepsilon_t,}$$

which is precisely the ADF regression form, and by part (a) the "level" coefficient ρ is the sum of the original AR coefficients.

Problem 3. 1. **Impact multiplier.**

Initial steady state (for $t \le 0$): $y_t = \bar{y}$ and $x_t = \bar{x}$. Plugging into the model,

$$\bar{y} = \alpha + \rho\bar{y} + (\beta_0 + \beta_1 + \beta_2)\bar{x} \quad \Longrightarrow \quad \bar{y} = \frac{\alpha + (\beta_0 + \beta_1 + \beta_2)\bar{x}}{1 - \rho}.$$

At $t = 1$ we have $x_1 = \bar{x} + 1$, while $y_0 = \bar{y}$, $x_0 = \bar{x}$, $x_{-1} = \bar{x}$. Then

$$y_1 = \alpha + \rho\bar{y} + \beta_0(\bar{x} + 1) + \beta_1\bar{x} + \beta_2\bar{x} + \varepsilon_1.$$

Compare this with what y_1 would be without the shock (i.e. with $x_1 = \bar{x}$); the only term that changes is $\beta_0 x_1$. Hence

$$\frac{\partial y_1}{\partial x_1} = \beta_0.$$

2. **Dynamic multipliers for $t = 2, 3, 4$.**

From part (a),

$$y_1 = \bar{y} + \beta_0 + (\varepsilon_1 - \bar{\varepsilon}),$$

but since we are looking at deterministic effects of the x-shock, take $\varepsilon_t = 0$ throughout and work with deviations from $\bar{y}$.

At $t = 2$: $x_2 = \bar{x}$ (shock is over), $x_1 = \bar{x} + 1$, $x_0 = \bar{x}$, and $y_1 = \bar{y} + \beta_0$. Then

$$y_2 = \alpha + \rho y_1 + \beta_0 x_2 + \beta_1 x_1 + \beta_2 x_0 = \alpha + \rho(\bar{y} + \beta_0) + \beta_0\bar{x} + \beta_1(\bar{x} + 1) + \beta_2\bar{x}.$$

Subtract the steady state value $\bar{y} = \alpha + \rho\bar{y} + (\beta_0 + \beta_1 + \beta_2)\bar{x}$:

$$y_2 - \bar{y} = \rho\beta_0 + \beta_1.$$

At $t = 3$: now $x_3 = \bar{x}$, $x_2 = \bar{x}$, $x_1 = \bar{x} + 1$, and $y_2 = \bar{y} + \rho\beta_0 + \beta_1$. Then

$$y_3 = \alpha + \rho y_2 + \beta_0\bar{x} + \beta_1\bar{x} + \beta_2(\bar{x} + 1).$$

Subtracting $\bar{y}$ again,

$$y_3 - \bar{y} = \rho(y_2 - \bar{y}) + \beta_2 = \rho(\rho\beta_0 + \beta_1) + \beta_2 = \rho^2\beta_0 + \rho\beta_1 + \beta_2.$$

At $t = 4$: all $x_4, x_3, x_2 = \bar{x}$. We have

$$y_4 = \alpha + \rho y_3 + \beta_0\bar{x} + \beta_1\bar{x} + \beta_2\bar{x},$$

so

$$y_4 - \bar{y} = \rho(y_3 - \bar{y}) = \rho(\rho^2\beta_0 + \rho\beta_1 + \beta_2) = \rho^3\beta_0 + \rho^2\beta_1 + \rho\beta_2.$$

General pattern (for $t \ge 3$):

$$y_t - \bar{y} = \rho^{t-1}\beta_0 \ + \ \rho^{t-2}\beta_1 \ + \ \rho^{t-3}\beta_2.$$

That is, the shock to x_1 propagates through the AR part via powers of ρ, while the three β's line up with lags of ρ.

3. **Long-run multiplier for a permanent increase.**

   Now suppose $x_t = \bar{x} + 1$ for all $t \geq 1$. In the new steady state we have

$$y_t = y_{t-1} = y^*, \qquad x_t = x_{t-1} = x_{t-2} = \bar{x} + 1.$$

   Plug into the model:

$$y^* = \alpha + \rho y^* + \beta_0(\bar{x}+1) + \beta_1(\bar{x}+1) + \beta_2(\bar{x}+1).$$

   Collect terms:

$$y^* - \rho y^* = \alpha + (\beta_0 + \beta_1 + \beta_2)(\bar{x}+1) \quad \Longrightarrow \quad (1-\rho)y^* = \alpha + (\beta_0 + \beta_1 + \beta_2)(\bar{x}+1).$$

   Hence

$$\boxed{y^* = \frac{\alpha + (\beta_0 + \beta_1 + \beta_2)(\bar{x}+1)}{1-\rho}}.$$

   If we differentiate the new steady state with respect to the permanent level of x (i.e. treat "+1" as "$+dx$"), we get the long-run multiplier:

$$\boxed{\frac{dy^*}{dx} = \frac{\beta_0 + \beta_1 + \beta_2}{1-\rho}}.$$

4. **General AR-DL(1,q) case.**

   For

$$y_t = \alpha + \rho y_{t-1} + \sum_{j=0}^{q} \beta_j x_{t-j} + \varepsilon_t,$$

   a permanent increase $x_t = \bar{x} + 1$ for all $t \geq 1$ implies in the new steady state

$$y^* = \alpha + \rho y^* + \sum_{j=0}^{q} \beta_j (\bar{x}+1).$$

   Thus

$$(1-\rho)y^* = \alpha + \Big(\sum_{j=0}^{q} \beta_j\Big)(\bar{x}+1), \quad \Longrightarrow \quad y^* = \frac{\alpha + \big(\sum_{j=0}^{q} \beta_j\big)(\bar{x}+1)}{1-\rho}.$$

   So the long-run effect of raising x by 1 permanently is

$$\boxed{\text{LRM} = \frac{\sum_{j=0}^{q} \beta_j}{1-\rho}.}$$

   Interpretation. The numerator $\sum_{j=0}^{q} \beta_j$ is the total contemporaneous + distributed-lag effect of x on y *before* accounting for feedback through y_{t-1}. Because current y feeds into future y through the AR term, every initial effect is propagated forward and summed as a geometric series with ratio ρ. If $\rho > 0$, these feedbacks reinforce the initial effect, magnifying it by $1/(1-\rho)$. If $\rho < 0$, the feedback oscillates in sign and the net long-run effect is dampened relative to the positive-lag sum.

Problem 4. We have the linear model

$$Y_t = X_t'\beta + e_t, \qquad \mathbb{E}[Z_t e_t] = 0, \qquad Z_t \in \mathbb{R}^\ell,\ X_t \in \mathbb{R}^k,\ \ell \geq k.$$

Let

$$\hat{m}_T(\beta) = \frac{1}{T}\sum_{t=1}^{T} Z_t(Y_t - X_t'\beta), \qquad \hat{S} \to_p S,$$

and the GMM estimator minimizes $\hat{m}_T(\beta)'\hat{S}^{-1}\hat{m}_T(\beta)$. The J-statistic is

$$J = T\hat{m}_T(\hat{\beta})'\hat{S}^{-1}\hat{m}_T(\hat{\beta}).$$

We prove step by step that $J \xrightarrow{d} \chi^2_{\ell-k}$.

1. Since S is the asymptotic variance of the moments, assume it is symmetric positive definite. Then we can take a Cholesky (or any square-root) factorization

   $$S^{-1} = CC', \qquad \text{with } C \text{ invertible.}$$

   Equivalently,

   $$S = C'^{-1}C^{-1},$$

   because premultiplying and postmultiplying by the inverses gives $(C'^{-1}C^{-1})^{-1} = CC' = S^{-1}$.

2. Start from

   $$J = T\hat{m}_T(\hat{\beta})'\hat{S}^{-1}\hat{m}_T(\hat{\beta}).$$

   Insert $I = C'^{-1}C'$ on the left and $I = CC^{-1}$ on the right, and use $\hat{S}^{-1} \approx CC'$:

   $$J = T\big(C'\hat{m}_T(\hat{\beta})\big)'\big(C'\hat{S}C\big)^{-1}\big(C'\hat{m}_T(\hat{\beta})\big).$$

   This is the desired expression.

3. Write the sample moment as

   $$\hat{m}_T(\beta) = \frac{1}{T}Z'(Y - X\beta), \quad \text{where } Z := (Z_1, \ldots, Z_T)',\ X := (X_1, \ldots, X_T)',\ Y := (Y_1, \ldots, Y_T)'.$$

   The linear GMM estimator $\hat{\beta}$ with weight $\hat{S}^{-1}$ solves the sample FOC

   $$\Big(\tfrac{1}{T}X'Z\Big)\hat{S}^{-1}\hat{m}_T(\hat{\beta}) = 0.$$

   Now write

   $$\hat{m}_T(\hat{\beta}) = \hat{m}_T(\beta) - \Big(\tfrac{1}{T}Z'X\Big)(\hat{\beta} - \beta),$$

   because $Y - X\hat{\beta} = (Y - X\beta) - X(\hat{\beta} - \beta)$. Plug this into the FOC:

   $$\Big(\tfrac{1}{T}X'Z\Big)\hat{S}^{-1}\Big[\hat{m}_T(\beta) - \Big(\tfrac{1}{T}Z'X\Big)(\hat{\beta} - \beta)\Big] = 0,$$

   so

   $$\Big(\tfrac{1}{T}X'Z\Big)\hat{S}^{-1}\hat{m}_T(\beta) = \Big(\tfrac{1}{T}X'Z\Big)\hat{S}^{-1}\Big(\tfrac{1}{T}Z'X\Big)(\hat{\beta} - \beta).$$

Hence
$$\hat\beta-\beta=\left[\Big(\tfrac{1}{T}X'Z\Big)\hat S^{-1}\Big(\tfrac{1}{T}Z'X\Big)\right]^{-1}\Big(\tfrac{1}{T}X'Z\Big)\hat S^{-1}\hat m_T(\beta).$$
Substitute this back into
$$\hat m_T(\hat\beta)=\hat m_T(\beta)-\Big(\tfrac{1}{T}Z'X\Big)(\hat\beta-\beta)$$
to get
$$\hat m_T(\hat\beta)=\hat m_T(\beta)-\Big(\tfrac{1}{T}Z'X\Big)\left[\Big(\tfrac{1}{T}X'Z\Big)\hat S^{-1}\Big(\tfrac{1}{T}Z'X\Big)\right]^{-1}\Big(\tfrac{1}{T}X'Z\Big)\hat S^{-1}\hat m_T(\beta).$$
Now premultiply by C':
$$C'\hat m_T(\hat\beta)=\left\{I_\ell-C'\Big(\tfrac{1}{T}Z'X\Big)\left[\Big(\tfrac{1}{T}X'Z\Big)\hat S^{-1}\Big(\tfrac{1}{T}Z'X\Big)\right]^{-1}\Big(\tfrac{1}{T}X'Z\Big)\hat S^{-1}C'^{-1}\right\}C'\hat m_T(\beta).$$
Define
$$A_T=I_\ell-C'\Big(\tfrac{1}{T}Z'X\Big)\left[\Big(\tfrac{1}{T}X'Z\Big)\hat S^{-1}\Big(\tfrac{1}{T}Z'X\Big)\right]^{-1}\Big(\tfrac{1}{T}X'Z\Big)\hat S^{-1}C'^{-1},$$
so that
$$C'\hat m_T(\hat\beta)=A_T\,C'\hat m_T(\beta).$$

4. By LLN,
$$\tfrac{1}{T}Z'X\xrightarrow{p}\mathbb{E}[Z_tX_t'],\qquad \tfrac{1}{T}X'Z\xrightarrow{p}\mathbb{E}[X_tZ_t']=\mathbb{E}[Z_tX_t']',\qquad \hat S\xrightarrow{p}S.$$
Hence
$$A_T\xrightarrow{p}I_\ell-C'\mathbb{E}[Z_tX_t']\big(\mathbb{E}[X_tZ_t']S^{-1}\mathbb{E}[Z_tX_t']\big)^{-1}\mathbb{E}[X_tZ_t']S^{-1}C'^{-1}.$$
Define
$$R:=C'\mathbb{E}[Z_tX_t']\in\mathbb{R}^{\ell\times k}.$$
Note that
$$\mathbb{E}[X_tZ_t']S^{-1}\mathbb{E}[Z_tX_t']=\big(\mathbb{E}[Z_tX_t']\big)'S^{-1}\big(\mathbb{E}[Z_tX_t']\big)=R'R,$$
because $S^{-1}=CC'$ and $R=C'\mathbb{E}[Z_tX_t']$. Also,
$$\mathbb{E}[X_tZ_t']S^{-1}C'^{-1}=\big(\mathbb{E}[Z_tX_t']\big)'CC'C'^{-1}=\big(\mathbb{E}[Z_tX_t']\big)'C=R'.$$
Therefore the probability limit is
$$A_T\xrightarrow{p}I_\ell-R(R'R)^{-1}R'.$$
This is the orthogonal projector onto the orthogonal complement of the column space of R.

5. By the multivariate CLT for i.i.d. (or suitable mixing) data,
$$T^{1/2}\hat m_T(\beta)=T^{1/2}\Big(\frac{1}{T}\sum_{t=1}^{T}Z_te_t\Big)\xrightarrow{d}N(0,S).$$
Premultiplying by C' and using $S=C'^{-1}C^{-1}$,
$$T^{1/2}C'\hat m_T(\beta)\xrightarrow{d}N(0,\,C'SC)=N(0,\,C'C'^{-1}C^{-1}C)=N(0,I_\ell).$$
So define
$$u:=\lim_d T^{1/2}C'\hat m_T(\beta)\sim N(0,I_\ell).$$

6. Recall from (b) and (c):

$$J = T\big(C'\hat{m}_T(\hat{\beta})\big)'\big(C'\hat{S}C\big)^{-1}\big(C'\hat{m}_T(\hat{\beta})\big).$$

But $C'\hat{S}C \xrightarrow{p} C'SC = I_\ell$, so

$$J = \big(T^{1/2}C'\hat{m}_T(\hat{\beta})\big)'\big(C'\hat{S}C\big)^{-1}\big(T^{1/2}C'\hat{m}_T(\hat{\beta})\big) \xrightarrow{d} \big(\lim T^{1/2}C'\hat{m}_T(\hat{\beta})\big)'\big(\lim C'\hat{S}C\big)^{-1}\big(\lim T^{1/2}C'\hat{m}_T(\hat{\beta})\big).$$

From (c) and (d),

$$T^{1/2}C'\hat{m}_T(\hat{\beta}) = A_T\, T^{1/2}C'\hat{m}_T(\beta) \xrightarrow{d} \big(I_\ell - R(R'R)^{-1}R'\big)u.$$

Also, $(C'\hat{S}C)^{-1} \to I_\ell$. Hence

$$J \xrightarrow{d} u'\big(I_\ell - R(R'R)^{-1}R'\big)u.$$

7. Finally, $P := I_\ell - R(R'R)^{-1}R'$ is the symmetric idempotent matrix that projects onto the orthogonal complement of $\operatorname{span}(R)$. Its rank is

$$\operatorname{rank}(P) = \ell - \operatorname{rank}(R).$$

Since $R = C'\mathbb{E}[Z_tX_t']$ is $\ell \times k$ and we assume the usual rank condition for GMM/IV (i.e. $\mathbb{E}[Z_tX_t']$ has full column rank k), we have $\operatorname{rank}(R) = k$. Thus

$$\operatorname{rank}(P) = \ell - k.$$

If $u \sim N(0, I_\ell)$ and P is a symmetric idempotent matrix of rank r, then

$$u'Pu \sim \chi^2_r.$$

Here $r = \ell - k$, so

$$u'\big(I_\ell - R(R'R)^{-1}R'\big)u \sim \chi^2_{\ell-k}.$$

Combining with (f) gives

$$J \xrightarrow{d} \chi^2_{\ell-k}.$$

Problem 5. We observe i.i.d. data $\{(Y_i, X_i)\}_{i=1}^n$ and want to estimate

$$\mu = \mathbb{E}[Y_i]$$

given the extra (true) moment

$$\mathbb{E}[X_i] = 0.$$

Consider the moment conditions

$$g_i(\mu) = \begin{bmatrix} Y_i - \mu \\ X_i \end{bmatrix}, \qquad \mathbb{E}[g_i(\mu)] = 0.$$

Then the sample moments are

$$\hat{g}_n(\mu) = \begin{bmatrix} \bar{Y} - \mu \\ \bar{X} \end{bmatrix}, \quad \bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i, \quad \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i.$$

Efficient GMM. Let

$$S = \mathrm{Var}\left(\begin{bmatrix} Y_i - \mu \\ X_i \end{bmatrix}\right) = \begin{bmatrix} \sigma_Y^2 & \sigma_{YX} \\ \sigma_{YX} & \sigma_X^2 \end{bmatrix},$$

where $\sigma_Y^2 = \mathrm{Var}(Y_i)$, $\sigma_X^2 = \mathrm{Var}(X_i)$, and $\sigma_{YX} = (Y_i, X_i)$. The efficient GMM weighting matrix is $W = S^{-1}$. The GMM estimator solves

$$\hat{\mu} = \arg\min_{\mu} \ \hat{g}_n(\mu)' W \hat{g}_n(\mu).$$

Write $W = (w_{jk})$. Then the FOC is

$$\frac{\partial}{\partial \mu}\big[(\bar{Y} - \mu, \bar{X}) W (\bar{Y} - \mu, \bar{X})'\big] = -2\big[w_{11}(\bar{Y} - \mu) + w_{12}\bar{X}\big] = 0,$$

so

$$w_{11}(\bar{Y} - \hat{\mu}) + w_{12}\bar{X} = 0 \quad \Longrightarrow \quad \hat{\mu} = \bar{Y} + \frac{w_{12}}{w_{11}}\bar{X}.$$

Now compute $W = S^{-1}$. Since

$$S^{-1} = \frac{1}{\sigma_Y^2\sigma_X^2 - \sigma_{YX}^2}\begin{bmatrix} \sigma_X^2 & -\sigma_{YX} \\ -\sigma_{YX} & \sigma_Y^2 \end{bmatrix},$$

we have

$$w_{11} = \frac{\sigma_X^2}{\sigma_Y^2\sigma_X^2 - \sigma_{YX}^2}, \qquad w_{12} = \frac{-\sigma_{YX}}{\sigma_Y^2\sigma_X^2 - \sigma_{YX}^2}.$$

Hence

$$\frac{w_{12}}{w_{11}} = -\frac{\sigma_{YX}}{\sigma_X^2}.$$

So the efficient GMM estimator is

$$\boxed{\hat{\mu} = \bar{Y} - \frac{(Y_i, X_i)}{\mathrm{Var}(X_i)}\,\bar{X},}$$

with the covariance and variance replaced by their sample analogs:

$$(Y_i, X_i) \approx \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X}), \qquad \mathrm{Var}(X_i) \approx \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Why does X_i help? If X_i is correlated with Y_i, then Y_i contains a component that is "predictable" from X_i. Because we *know* the true mean of X_i is 0, we can remove from $\bar{Y}$ the part that is due to a nonzero sample mean $\bar{X}$. The correction

$$-\frac{(Y, X)}{\mathrm{Var}(X)}\bar{X}$$

is exactly the control–variate adjustment: it has mean zero but lowers the variance of the estimator. If X_i is uncorrelated with Y_i ($(Y, X) = 0$), the adjustment is zero and we are back to $\hat{\mu} = \bar{Y}$.

Problem 6. We observe $\{(Y_i, X_i, Z_i)\}_{i=1}^n$ and the econometrician posits the linear IV/GMM model

$$Y_i = X_i'\beta + e_i, \qquad \mathbb{E}[Z_i e_i] = 0,$$

with $\ell = \dim Z_i > k = \dim X_i > 1$. We study what happens when this moment is (locally) misspecified.

1. **GMM estimator with a given weight matrix $W > 0$.**

   Define the sample moment

   $$\hat{m}_n(\beta) := \frac{1}{n}\sum_{i=1}^n Z_i(Y_i - X_i'\beta) = \frac{1}{n}Z'(Y - X\beta),$$

   where $Z = (Z_1', \ldots, Z_n')'$, $X = (X_1', \ldots, X_n')'$, $Y = (Y_1, \ldots, Y_n)'$. Given a fixed positive definite weight matrix W (of size $\ell \times \ell$), the GMM estimator is

   $$\hat{\beta} = \arg\min_{\beta} \ \hat{m}_n(\beta)' W \hat{m}_n(\beta).$$

   For the linear case this has the closed form

   $$\boxed{\hat{\beta} = \big((n^{-1}X'Z)W(n^{-1}Z'X)\big)^{-1}(n^{-1}X'Z)W(n^{-1}Z'Y).}$$

2. **Misspecification implies $\mathbb{E}[Ze] \neq 0$.**

   We are told the true error is

   $$e_i = \frac{\delta}{\sqrt{n}} + u_i, \qquad \mathbb{E}[u_i \mid Z_i] = 0,$$

   and $\mu_Z := \mathbb{E}[Z_i] \neq 0$, with $\delta \neq 0$. Then

   $$\mathbb{E}[Z_i e_i] = \mathbb{E}\big[Z_i(\delta/\sqrt{n} + u_i)\big] = \frac{\delta}{\sqrt{n}}\mathbb{E}[Z_i] + \mathbb{E}[Z_i u_i].$$

   Because $\mathbb{E}[u_i \mid Z_i] = 0 \Rightarrow \mathbb{E}[Z_i u_i] = 0$, we get

   $$\mathbb{E}[Z_i e_i] = \frac{\delta}{\sqrt{n}}\mu_Z \neq 0$$

   when $\delta \neq 0$ and $\mu_Z \neq 0$. So the moment $\mathbb{E}[Ze] = 0$ is violated at order $1/\sqrt{n}$.

3. **Linear expansion of $\sqrt{n}(\hat{\beta} - \beta)$.**

   Start from the sample FOC for linear GMM:

   $$(n^{-1}X'Z)W\hat{m}_n(\hat{\beta}) = 0.$$

   Write the sample moment at $\hat{\beta}$ as

   $$\hat{m}_n(\hat{\beta}) = \hat{m}_n(\beta) - (n^{-1}Z'X)(\hat{\beta} - \beta),$$

because $Y - X\hat{\beta} = (Y - X\beta) - X(\hat{\beta} - \beta)$. Plugging in,

$$(n^{-1}X'Z)W\big[\hat{m}_n(\beta) - (n^{-1}Z'X)(\hat{\beta} - \beta)\big] = 0,$$

so

$$(n^{-1}X'Z)W\hat{m}_n(\beta) = (n^{-1}X'Z)W(n^{-1}Z'X)(\hat{\beta} - \beta).$$

Hence

$$\hat{\beta} - \beta = \big[(n^{-1}X'Z)W(n^{-1}Z'X)\big]^{-1}(n^{-1}X'Z)W\hat{m}_n(\beta).$$

Now substitute the misspecified error:

$$\hat{m}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} Z_i e_i = \frac{1}{n}\sum_{i=1}^{n} Z_i\Big(\frac{\delta}{\sqrt{n}} + u_i\Big) = \frac{\delta}{\sqrt{n}}\Big(\frac{1}{n}\sum_{i=1}^{n} Z_i\Big) + \frac{1}{n}\sum_{i=1}^{n} Z_i u_i.$$

Therefore

$$\sqrt{n}(\hat{\beta} - \beta) = \big[(n^{-1}X'Z)W(n^{-1}Z'X)\big]^{-1}(n^{-1}X'Z)W\Big[\delta\Big(\frac{1}{n}\sum_{i=1}^{n} Z_i\Big) + \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n} Z_i u_i\Big)\Big].$$

This is the requested expression in terms of W, n, δ and (X_i, Z_i, u_i).

4. **Asymptotic distribution and asymptotic bias.**

   Let

$$G := \mathbb{E}[Z_i X_i'] \quad (\ell \times k), \qquad \mu_Z := \mathbb{E}[Z_i], \qquad \Omega := \mathrm{Var}(Z_i u_i) \quad (\ell \times \ell).$$

   By LLN/CLT,

$$n^{-1}X'Z \xrightarrow{p} G', \qquad n^{-1}Z'X \xrightarrow{p} G, \qquad \frac{1}{n}\sum_{i=1}^{n} Z_i \xrightarrow{p} \mu_Z,$$

   and

$$\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n} Z_i u_i\Big) \xrightarrow{d} N(0, \Omega).$$

   Define

$$Q := G'WG \quad (k \times k), \qquad \text{nonsingular by relevance.}$$

   Then, by continuous mapping,

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} Q^{-1}G'W\big(\delta\mu_Z + \eta\big), \quad \eta \sim N(0, \Omega).$$

   So

$$\sqrt{n}(\hat{\beta} - \beta) \;\xrightarrow{d}\; N\Big(\underbrace{Q^{-1}G'W\delta\mu_Z}_{\text{asymptotic bias}},\; Q^{-1}G'W\Omega WGQ^{-1}\Big).$$

   The mean is nonzero whenever $\delta \neq 0$ and $\mu_Z \neq 0$, so the limiting distribution is centered away from 0: this is the asymptotic bias induced by the local misspecification.

5. **Does the optimal weight remove the bias?**

   The optimal weight is $W = \Omega^{-1}$. Plugging into the bias term,

$$\text{bias} = (G'\Omega^{-1}G)^{-1}G'\Omega^{-1}\delta\mu_Z.$$

   This is generally *not* zero unless μ_Z happens to lie in the null space of $G'\Omega^{-1}$ (a non-generic case). So choosing the optimal weight minimizes asymptotic variance but does *not* remove the bias created by the drifting violation $\mathbb{E}[Ze] = (\delta/\sqrt{n})\mu_Z$.

6. **Does this misspecification affect consistency?**

   Here the violation is *local*:

$$\mathbb{E}[Z_i e_i] = \frac{\delta}{\sqrt{n}}\mu_Z \;\longrightarrow\; 0 \quad \text{as } n \to \infty.$$

   So the population moment condition is satisfied in the limit. Therefore the GMM estimator still converges in probability to the true β:

$$\hat{\beta} \xrightarrow{p} \beta.$$

   What changes is the *center* of the $\sqrt{n}$-distribution: it is no longer 0 but the nonzero vector in part (d). So: the misspecification does *not* break consistency, but it does create an asymptotic bias at the $\sqrt{n}$ scale.