Large Sample Theory – Analytical Exercises

Ryan Benschop

1.

Claim: Define the stochastic process $\{z_n\}_{n\in\mathbb{N}}$ *by:*

$$\begin{aligned}\mathbb{P}(z_n = 0) &= 1 - n^{-1}\\ \mathbb{P}(z_n = n^2) &= n^{-1}\end{aligned}$$

Then, $z_n \xrightarrow{p} 0$ *and* $\mathbb{E}z_n \to \infty$*.*

Proof. To show that $z_n \xrightarrow{p} 0$, we need to show that for any $\varepsilon, \delta \in \mathbb{R}_{++}$, there is $N \in \mathbb{N}$ such that, if $n \geq N$, $\mathbb{P}(|z_n| > \varepsilon) < \delta$.

Choose any $\varepsilon, \delta \in \mathbb{R}_{++}$. Let $N \in \mathbb{N}$ be such that $N > \delta^{-1}$. Notice that the event $|z_n| > \varepsilon$ is a subset of the complementary event of $z_n = 0$. Hence:

$$\begin{aligned}\mathbb{P}(|z_n| > \varepsilon) &\leq 1 - \mathbb{P}(z_n = 0)\\ &= 1 - (1 - n^{-1})\\ &= n^{-1}\\ &< \delta\end{aligned}$$

Next, to show that $\mathbb{E}z_n \to \infty$, we need to show that, for any $\Delta \in \mathbb{R}$, there is $N \in \mathbb{N}$ such that, if $n \geq N$, $\mathbb{E}z_n > \Delta$.

Choose any $\Delta \in \mathbb{R}$. Let $N \in \mathbb{N}$ be such that $N > \Delta$. Notice, $\mathbb{E}(z_n) = n$. Hence, if $n \geq N$, $\mathbb{E}(z_n) > \Delta$. □

2. (Chebychev's weak law of large numbers.)

Claim: *Let* $\{z_n\}_{n\in\mathbb{N}}$ *be a stochastic process. Suppose* $\mathbb{E}\bar{z}_n = \mu$ *and* $\mathbb{V}z_n = 0$*. Then,* $\bar{z}_n \xrightarrow{p} \mu$*.*

Proof. To prove the theorem, we can show that under the given assumptions, $\bar{z}_n \xrightarrow{m.s.} \mu$, noting that convergence in mean square imples convergence in probability. We need to show that $\mathbb{E}[(z_n - \mu)^2] \to 0$.

Note:

$$\begin{aligned}\mathbb{E}[(z_n - \mu)^2] &= \mathbb{E}[(\bar{z}_n - \mathbb{E}[\bar{z}_n] + \mathbb{E}[\bar{z}_n] - \mu)^2] \\ &= \mathbb{E}[(\bar{z}_n - \mathbb{E}[\bar{z}_n])^2 - 2(\bar{z}_n - \mathbb{E}[\bar{z}_n])(\mathbb{E}[\bar{z}_n] - \mu) + (\mathbb{E}[\bar{z}_n] - \mu)^2]\end{aligned}$$

Next, note:

$$\begin{aligned}(\bar{z}_n - \mathbb{E}[\bar{z}_n])(\mathbb{E}[\bar{z}_n] - \mu) &= \mathbb{E}[\bar{z}_n\mathbb{E}[\bar{z}_n] - \bar{z}_n\mu - (\mathbb{E}[\bar{z}_n])^2 + \mathbb{E}[\bar{z}_n]\mu] \\ &= (\mathbb{E}[\bar{z}_n])^2 - \mathbb{E}[\bar{z}_n]\mu - (\mathbb{E}[\bar{z}_n])^2 + \mathbb{E}[\bar{z}_n]\mu \\ &= 0\end{aligned}$$

So:

$$\begin{aligned}\mathbb{E}[(z_n - \mu)^2] &= \mathbb{E}[(\bar{z}_n - \mathbb{E}[\bar{z}_n])^2 + (\mathbb{E}[\bar{z}_n] - \mu)^2] \\ &= \mathbb{E}[(\bar{z}_n - \mathbb{E}[\bar{z}_n])^2] + \mathbb{E}[(\mathbb{E}[\bar{z}_n] - \mu)^2] \\ &= \mathbb{V}[\bar{z}_n] + (\mathbb{E}[\bar{z}_n])^2 - 2\mathbb{E}[\bar{z}_n]\mu + \mu^2 \\ &\to 0 + \mu^2 - 2\mu^2 + \mu^2 = 0\end{aligned}$$

□

3. (Consistency and asymptotic normality of OLS for random samples)

Claim: Consider the stochastic process $\{(y_i, x_i)\}_{i\in\mathbb{N}}$ *where, for each* $i \in \mathbb{N}$, y_i *is* $\mathbb{R}$*-valued, and* x_i *is* $\mathbb{R}^K$*-valued. Suppose the following assumptions hold:*

- *(Linearity): there is* $\beta \in \mathbb{R}^K$ *such that, for each* $i \in \mathbb{N}$, $y_i = x_i'\beta + \varepsilon_i$.
- *(Random sampling):* $\{(y_i, x_i)\}_{i\in\mathbb{N}}$ *is an i.i.d. process.*
- *(Predetermined regressors): for each* $i \in \mathbb{N}$, $\mathbb{E}[x_i\varepsilon_i] = 0$.
- *(Rank condition):* $\mathbb{E}[x_i x_i']$ *exists and is nonsingular.*

Let $\hat{\beta}$ *denote the OLS estimator of* β*. Then,* $\hat{\beta}$ *is consistent. Suppose, further, that* $\mathbb{E}[\varepsilon_i^2 x_i x_i']$ *exists and is finite. Then, the asymptotic distribution related to* $\hat{\beta}$ *is given by:*

$$n^{\frac{1}{2}}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, Asy.\mathbb{V}[\hat{\beta}])$$

where:

$$Asy.\mathbb{V}[\hat{\beta}] = (\mathbb{E}[x_i x_i'])^{-1}\mathbb{E}[\varepsilon_i^2 x_i x_i'](\mathbb{E}[x_i x_i'])^{-1}$$

Proof. (Direct proof) Letting X denote the data matrix, and rewriting the expression for the OLS estimator, we have:

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + \varepsilon)\\ &= \beta + (X'X)^{-1}X'\varepsilon\\ &= \beta + \left(n^{-1}\sum_{i=1}^n x_i x_i'\right)^{-1} n^{-1}\sum_{i=1}^n x_i\varepsilon_i\end{aligned}$$

Now, notice that, since $\{(y_i, x_i)\}_{i\in\mathbb{N}}$ is i.i.d., $\{x_i x_i'\}_{i\in\mathbb{N}}$ and $\{x_i\varepsilon_i\}_{i\in\mathbb{N}}$ are also i.i.d. (stochastic processes defined by measurable functions of i.i.d. processes are i.i.d.). Additionally, their means exist by assumption. Thus, by Kolmogorov's SLLN:

$$n^{-1}\sum_{i=1}^n x_i x_i' \xrightarrow{a.s.} \mathbb{E}[x_i x_i'] \qquad \text{and} \qquad n^{-1}\sum_{i=1}^n x_i\varepsilon_i \xrightarrow{a.s.} \mathbb{E}[x_i\varepsilon_i]$$

which implies that convergence in probability holds as well. Define $g : \mathbb{R}^{K,K} \times \mathbb{R}^{K,1} \to \mathbb{R}^K$ by:

$$g(A, B) = \beta + A^{-1}B$$

Since we assumed $\mathbb{E}[x_i x_i']$ is invertible, noting that matrix inversion is a continuous transformation, the continuous mapping theorem implies:

$$\begin{aligned}\operatorname*{plim}_{n\to\infty}\hat{\beta} &= \operatorname*{plim}_{n\to\infty} g\left(n^{-1}\sum_{i=1}^n x_i x_i', n^{-1}\sum_{i=1}^n x_i\varepsilon_i\right) = g\left(\operatorname*{plim}_{n\to\infty} n^{-1}\sum_{i=1}^n x_i x_i', \operatorname*{plim}_{n\to\infty} n^{-1}\sum_{i=1}^n x_i\varepsilon_i\right)\\ &= g(\mathbb{E}[x_i x_i'], \mathbb{E}[x_i\varepsilon_i]) = \beta + \mathbb{E}[x_i x_i']^{-1}\mathbb{E}[x_i\varepsilon_i]\end{aligned}$$

But, by assumption, $\mathbb{E}[x_i\varepsilon_i] = 0$. Hence, $\hat{\beta} \xrightarrow{p} \beta$.

Now, rearranging the above expression for $\hat{\beta}$, we have:

$$n^{\frac{1}{2}}(\hat{\beta}-\beta) = \left(n^{-1}\sum_{i=1}^{n} x_i x_i'\right)^{-1} n^{-\frac{1}{2}}\sum_{i=1}^{n} x_i\varepsilon_i$$

Recalling that $\{x_i\varepsilon_i\}_{i\in\mathbb{N}}$ is i.i.d., noting that its mean is zero and its variance exists by assumption, the Lindeberg-Levy central limit theorem implies:

$$n^{-\frac{1}{2}}\sum_{i=1}^{n} x_i\varepsilon_i \xrightarrow{d} \mathcal{N}(0,\Omega)$$

where $\Omega = \mathbb{V}[x_i\varepsilon_i] = \mathbb{E}[\varepsilon_i^2 x_i x_i']$.

Thus, by the continuous mapping theorem:

$$n^{-\frac{1}{2}}(\hat{\beta}-\beta) \xrightarrow{d} (\mathbb{E}[x_i x_i])^{-1}\Xi$$

where $\Xi \sim \mathcal{N}(0,\Omega)$. Hence:

$$n^{-\frac{1}{2}}(\hat{\beta}-\beta) \xrightarrow{d} \mathcal{N}(0,\text{Asy.}\mathbb{V}[\hat{\beta}])$$

noting that $(\mathbb{E}[x_i x_i'])^{-1}\Omega(\mathbb{E}[x_i x_i'])^{-1} = (\mathbb{E}[x_i x_i'])^{-1}\mathbb{E}[\varepsilon_i^2 x_i x_i'](\mathbb{E}[x_i x_i'])^{-1} = \text{Asy.}\mathbb{V}[\hat{\beta}]$ □

Proof. (Using Proposition 2.1 in the text) Proposition 2.1. provides that under the assumptions of linearity, ergodic stationarity of $\{(y_i,x_i)\}_{i\in\mathbb{N}}$, predetermined regressors, and satisfaction of the rank condition, the OLS estimator is consistent. Moreover, under the additional assumption that $\{x_i\varepsilon_i\}_{i\in\mathbb{N}}$ is a martingale difference sequence with finite second moments, then:

$$n^{\frac{1}{2}}(\hat{\beta}-\beta) \xrightarrow{d} \mathcal{N}(0,\text{Asy.}\mathbb{V}[\hat{\beta}])$$

Hence, we need only show that under the first set of assumptions in the present claim, $\{(y_i,x_i)\}_{i\in\mathbb{N}}$ is stationary ergodic and, under the further second moment assumption, $\{x_i\varepsilon_i'\}_{i\in\mathbb{N}}$ is a martingale difference sequence with finite second moments.

For $i\in\mathbb{N}$, let $z_i=(y_i,x_i)$. Let $i_1,\ldots,i_r$ be a set of indices, and fix any $i\in\mathbb{N}$ with $i\le i_1$, and $k\in\mathbb{N}$. Consider the joint distributions $(z_i,z_{i_1},\ldots,z_{i_r})$ and $(z_{i+k},z_{i_1+k},\ldots,z_{i_r+k})$. Since z_i is an independent process, each of these joint densities is simply the product of the marginal densities of the individual elements.[1] Further, since the elements are identically distributed, each of these marginal densities is the same. Hence, the two joint densities are the same, so $\{z_i\}_{i\in\mathbb{N}}$ is a stationary process.

Next, let $f:\mathbb{R}^K\to\mathbb{R}$ and $\mathbb{R}^\ell\to\mathbb{R}$ (for any $\ell\in\mathbb{N}$) be any bounded functions. Since z_i is an independent process:

$$\mathbb{E}[f(z_i,\ldots,z_{i+k})g(z_{i+n},\ldots,z_{i+n+\ell})] = \mathbb{E}[f(z_i,\ldots,z_{i+k})]\mathbb{E}[g(z_{i+n},\ldots,z_{i+n+\ell})]$$

for any $n\in\mathbb{N}$. Hence:

$$\lim_{n\to\infty}|\mathbb{E}[f(z_i,\ldots,z_{i+k})g(z_{i+n},\ldots,z_{i+n+\ell})]| = \mathbb{E}[f(z_i,\ldots,z_{i+k})]\mathbb{E}[g(z_{i+n},\ldots,z_{i+n+\ell})]$$

so that z_i is ergodic.

[1] Note that 'marginal density' here refers to the density of $z_i=(y_i,x_i)$.

Finally, to see that $\{x_i\varepsilon_i\}_{i\in\mathbb{N}}$ is a martingale difference sequence, notice that, since $\{(y_i, x_i)\}_{i\in\mathbb{N}}$ is i.i.d., $\{x_i\varepsilon_i\}_{i\in\mathbb{N}}$ is i.i.d. Thus, $\mathbb{E}[x_i\varepsilon_i|x_{i-1}\varepsilon_{i-1}, x_{i-2}\varepsilon_{i-2}, \ldots] = \mathbb{E}[x_i\varepsilon_i]$. But, by the assumption of predetermined regressors, this expectation is zero. Hence, $\{x_i\varepsilon_i\}_{i\in\mathbb{N}}$ is a martingale difference sequence. $\square$

4. (Consistent estimation of $\mathbb{E}(\varepsilon_i^2 x_i x_i')$)

Claim: Consider the $\mathbb{R}^2$-valued stochastic process $\{(y_i, x_i)\}_{i\in\mathbb{N}}$. Suppose:

- *For each $i \in \mathbb{N}$, $y_i = \beta x_i + \varepsilon_i$.*
- *$\{(y_i, x_i)\}$ is jointly stationary and ergodic.*
- *$\mathbb{E}[\varepsilon_i^2 x_i^2]$ exists and is finite.*
- *$\mathbb{E}[x_i^4]$ exists and is finite.*

Let $\hat{\beta}$ be a consistent estimator of β. Then:

$$n^{-1}\sum_{i=1}^{n} \hat{\varepsilon}_i^2 x_i^2 \xrightarrow{p} \mathbb{E}(\varepsilon_i^2 x_i^2)$$

Proof. Note:

$$\begin{aligned}
\hat{\varepsilon}_i &= y_i - \hat{\beta} x_i \\
&= y_i - \hat{\beta} x_i + \beta x_i - \beta x_i \\
&= y_i - \beta x_i - (\hat{\beta} - \beta) x_i \\
&= \varepsilon_i - (\hat{\beta} - \beta) x_i \\
\implies \hat{\varepsilon}_i^2 &= \varepsilon_i^2 - 2(\hat{\beta} - \beta)\varepsilon_i x_i + (\hat{\beta} - \beta)^2 x_i^2 \\
\implies \hat{\varepsilon}_i^2 x_i^2 &= \varepsilon_i^2 x_i^2 - 2(\hat{\beta} - \beta)\varepsilon_i x_i^3 + (\hat{\beta} - \beta)^2 x_i^4
\end{aligned}$$

Hence:

$$n^{-1}\sum_{i=1}^{n} \hat{\varepsilon}_i^2 x_i^2 = n^{-1}\sum_{i=1}^{n} \varepsilon_i^2 x_i^2 - 2(\hat{\beta} - \beta) n^{-1}\sum_{i=1}^{n} \varepsilon_i x_i^3 + (\hat{\beta} - \beta)^2 n^{-1}\sum_{i=1}^{n} x_i^4$$

Now, consider the convergence properties of these terms in turn. Since $\{y_i, x_i\}$ is jointly stationary and ergodic, $\{\varepsilon_i^2 x_i^2\}$ is stationary and ergodic. Moreover, the mean exists by assumption. Hence, by Kolmogorov's SLLN:

$$n^{-1}\sum_{i=1}^{n} \varepsilon_i^2 x_i^2 \xrightarrow{a.s.} \mathbb{E}[\varepsilon_i^2 x_i^2]$$

Next, consider the term $n^{-1}\sum_{i=1}^{n} \varepsilon_i x_i^3$. Note that $\{\varepsilon_i x_i^3\}$ is also stationary ergodic. Moreover, by the Cauchy-Schwarz inequality:

$$\mathbb{E}[|\varepsilon_i x_i^3|] = \mathbb{E}[|(\varepsilon_i x_i) x_i^2|] \le \left(\mathbb{E}[\varepsilon_i^2 x_i^2]\mathbb{E}[x_i^4]\right)^{\frac{1}{2}}$$

Since the means on the R.H.S. exist and are finite, $\mathbb{E}[\varepsilon_i x_i^3]$ does as well. Hence, by the SLLN:

$$n^{-1}\sum_{i=1}^{n} \varepsilon_i x_i^3 \xrightarrow{a.s.} \mathbb{E}[\varepsilon_i x_i^3]$$

Since $\hat{\beta}$ is consistent ($\hat{\beta} - \beta \xrightarrow{p} 0$), the continuous mapping theorem implies:

$$2(\hat{\beta} - \beta)n^{-1}\sum_{i=1}^{n}\varepsilon_i x_i^3 \xrightarrow{p} 0$$

Finally, consider $n^{-1}\sum_{i=1}^{n} x_i^4$. Given the ergodic stationarity and the fourth moment assumption, the SLLN implies:

$$n^{-1}\sum_{i=1}^{n} x_i^4 \xrightarrow{a.s.} \mathbb{E}[x_i^4]$$

Using the same argument as above, we have:

$$(\hat{\beta} - \beta)^2 n^{-1}\sum_{i=1}^{n} x_i^4 \xrightarrow{p} 0$$

Combining these convergence results, we have:

$$n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_i^2 x_i^2 \xrightarrow{p} \mathbb{E}[\varepsilon_i^2 x_i^2]$$

Note that the convergence is strong if the consistency of $\hat{\beta}$ is strong. $\square$

8. (Population analogue of deviation from the mean regression formula.)

Claim: *Let (y, x) be a random variable where y is $\mathbb{R}$-valued and x is $\mathbb{R}^K$-valued, with the first term of x being one. Let $x = (1, \tilde{x})$, so that $\tilde{x}$ contains the nonconstant covariates. Denote, by $\hat{\mathbb{E}}^*(y|x)$, the linear projection of y on x, assuming $\mathbb{E}[xx']$ is nonsingular. Then:*

$$\hat{\mathbb{E}}^*(y|x) = \mu + \gamma'\tilde{x}$$

where:

$$\begin{aligned}\gamma &= \mathbb{V}(\tilde{x})^{-1}\mathbb{C}(\tilde{x}, y)\\ \mu &= \mathbb{E}(y) - \mathbb{E}(\tilde{x})'\gamma\end{aligned}$$

Proof.

$$\begin{aligned}
\hat{\mathbb{E}}^*(y|x) &= x'(\mathbb{E}[xx'])^{-1}\mathbb{E}[xy]\\
&= \begin{bmatrix}1 & \tilde{x}'\end{bmatrix}\left(\mathbb{E}\begin{bmatrix}1 & \tilde{x}'\\ \tilde{x} & \tilde{x}\tilde{x}'\end{bmatrix}\right)^{-1}\mathbb{E}\begin{bmatrix}y\\ \tilde{x}y\end{bmatrix}\\
&= \begin{bmatrix}1 & \tilde{x}'\end{bmatrix}\begin{bmatrix}1 & \mathbb{E}[\tilde{x}']\\ \mathbb{E}[\tilde{x}] & \mathbb{E}[\tilde{x}\tilde{x}']\end{bmatrix}^{-1}\begin{bmatrix}\mathbb{E}[y]\\ \mathbb{E}[\tilde{x}y]\end{bmatrix}\\
&= \begin{bmatrix}1 & \tilde{x}'\end{bmatrix}\begin{bmatrix}1 + \mathbb{E}[\tilde{x}'](\mathbb{E}[\tilde{x}\tilde{x}'] - \mathbb{E}[\tilde{x}]\mathbb{E}[\tilde{x}'])^{-1}\mathbb{E}[\tilde{x}] & -\mathbb{E}[\tilde{x}'](\mathbb{E}[\tilde{x}\tilde{x}'] - \mathbb{E}[\tilde{x}]\mathbb{E}[\tilde{x}'])^{-1}\\ -(\mathbb{E}[\tilde{x}\tilde{x}'] - \mathbb{E}[\tilde{x}]\mathbb{E}[\tilde{x}'])^{-1}\mathbb{E}[\tilde{x}] & (\mathbb{E}[\tilde{x}\tilde{x}'] - \mathbb{E}[\tilde{x}]\mathbb{E}[\tilde{x}'])^{-1}\end{bmatrix}\begin{bmatrix}\mathbb{E}[y]\\ \mathbb{E}[\tilde{x}y]\end{bmatrix}\\
&= \begin{bmatrix}1 & \tilde{x}'\end{bmatrix}\begin{bmatrix}1 + \mathbb{E}[\tilde{x}']\mathbb{V}[\tilde{x}]^{-1}\mathbb{E}[\tilde{x}] & -\mathbb{E}[\tilde{x}']\mathbb{V}[\tilde{x}]^{-1}\\ -\mathbb{V}[\tilde{x}]^{-1}\mathbb{E}[\tilde{x}] & \mathbb{V}[\tilde{x}]^{-1}\end{bmatrix}\begin{bmatrix}\mathbb{E}[y]\\ \mathbb{E}[\tilde{x}y]\end{bmatrix}\\
&= \begin{bmatrix}1 + \mathbb{E}[\tilde{x}']\mathbb{V}[\tilde{x}]^{-1}\mathbb{E}[\tilde{x}] - \tilde{x}'\mathbb{V}[\tilde{x}]^{-1}\mathbb{E}[\tilde{x}] & -\mathbb{E}[\tilde{x}']\mathbb{V}[\tilde{x}]^{-1} + \tilde{x}'\mathbb{V}[\tilde{x}]^{-1}\end{bmatrix}\begin{bmatrix}\mathbb{E}[y]\\ \mathbb{E}[\tilde{x}y]\end{bmatrix}\\
&= \mathbb{E}[y] + \mathbb{E}[\tilde{x}']\mathbb{V}[\tilde{x}]^{-1}\mathbb{E}[\tilde{x}]\mathbb{E}[y] - \tilde{x}'\mathbb{V}[\tilde{x}]^{-1}\mathbb{E}[\tilde{x}]\mathbb{E}[y] - \mathbb{E}[\tilde{x}']\mathbb{V}[\tilde{x}]^{-1}\mathbb{E}[\tilde{x}y] + \tilde{x}'\mathbb{V}[\tilde{x}]^{-1}\mathbb{E}[\tilde{x}y]\\
&= \mathbb{E}[y] + \mathbb{E}[\tilde{x}']\Big(\mathbb{V}[\tilde{x}]^{-1}\mathbb{E}[\tilde{x}]\mathbb{E}[y] - \mathbb{V}[\tilde{x}]^{-1}\mathbb{E}[\tilde{x}y]\Big) + \tilde{x}'\Big(\mathbb{V}[\tilde{x}]^{-1}\mathbb{E}[\tilde{x}y] - \mathbb{V}[\tilde{x}]^{-1}\mathbb{E}[\tilde{x}]\mathbb{E}[y]\Big)\\
&= \mathbb{E}[y] - \mathbb{E}[\tilde{x}']\mathbb{V}[\tilde{x}']^{-1}\mathbb{C}[\tilde{x}, y] + \tilde{x}'\mathbb{V}[\tilde{x}]^{-1}\mathbb{C}[\tilde{x}, y]\\
&= \mu + \tilde{x}'\gamma
\end{aligned}$$

$\square$

11. (Breusch-Godfrey test for serial correlation)

Claim: Consider the standard regression framework. Let X denote the data matrix, and let:

$$E = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \hat{\varepsilon}_1 & 0 & \cdots & 0 \\ \hat{\varepsilon}_2 & \hat{\varepsilon}_1 & \cdots & 0 \\ \hat{\varepsilon}_3 & \hat{\varepsilon}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\varepsilon}_{n-1} & \hat{\varepsilon}_{n-2} & \cdots & \hat{\varepsilon}_{n-p} \end{bmatrix} \qquad \hat{B} = \begin{bmatrix} n^{-1}X'X & n^{-1}X'E \\ n^{-1}E'X & n^{-1}E'E \end{bmatrix} \qquad \hat{B}^{-1} = \begin{bmatrix} \hat{B}^{11} & \hat{B}^{12} \\ \hat{B}^{21} & \hat{B}^{22} \end{bmatrix}$$

$$\hat{\varepsilon} = \begin{bmatrix} \hat{\varepsilon}_1 & \cdots & \hat{\varepsilon}_n \end{bmatrix}' \qquad \hat{\gamma} = \begin{bmatrix} \hat{\gamma}_1 & \cdots & \hat{\gamma}_p \end{bmatrix}'$$

Then:

(a) In the auxiliary regression of $\hat{\varepsilon}_t$ on $(x_t, \hat{\varepsilon}_{t-1}, \ldots, \hat{\varepsilon}_{t-p})$:

$$\hat{\alpha} = \hat{B}^{-1} \begin{bmatrix} 0 \\ \hat{\gamma} \end{bmatrix}$$

where $\hat{\alpha}$ denotes the OLS estimator in the auxiliary regression.

(b) $\hat{B} \xrightarrow{p} B$, where:

$$B = \begin{bmatrix} \mathbb{E}(x_t x_t') & H \\ H' & \sigma^2 I \end{bmatrix}$$
$$H = \begin{bmatrix} \mathbb{E}(x_t \varepsilon_{t-1}) & \cdots & \mathbb{E}(x_t \varepsilon_{t-p}) \end{bmatrix}$$

(c) $\hat{\alpha} \xrightarrow{p} 0$.

(d)

$$\frac{SSR}{n-K-p} \xrightarrow{p} \sigma^2$$

where SSR denote the sum of squared residuals from the auxiliary regression.

(e)

$$pF = \frac{n\hat{\gamma}'\hat{B}^{22}\hat{\gamma}}{SSR/(n-K-p)}$$

where F denotes the F statistic corresponding with the hypothesis that the coefficients on the lagged residuals are equal to zero.

(f)

$$\hat{B}^{22} = \left[n^{-1}E'E - \left(n^{-1}E'X\right)n^{-1}X'X\left(n^{-1}X'E\right)\right]^{-1}$$

(g) $Q^*_{BP} - pF \xrightarrow{p} 0$, where:

$$Q^*_{BP} = n\hat{\rho}'(I - \hat{\Phi})^{-1}\hat{\rho}$$

is the modified Box-Pierce statistic used for testing whether the error term is autocorrelated, $\hat{\rho}$ denotes the vector of p sample autocorrelations of the residuals, and $\hat{\Phi}$ is defined elementwise by:

$$\hat{\phi}_{jk} = n^{-1} \sum_{t=j+1}^{n} x_t \hat{\varepsilon}_{t-j} \left(n^{-1} \sum_{t=1}^{n} x_t x_t' \right)^{-1} n^{-1} \sum_{t=k+1}^{n} x_t \hat{\varepsilon}_{t-k} \left([n-K]^{-1} \sum_{t=1}^{n} \hat{\varepsilon}_t^2 \right)^{-1}$$

Proof. (a): The data matrix corresponding to the auxiliary regression is: $\Xi = \begin{bmatrix} X & E \end{bmatrix}$. So:

$$\begin{aligned}
\hat{\alpha} &= (\Xi'\Xi)^{-1}\Xi'\hat{\varepsilon} \\
&= (n^{-1}\Xi'\Xi)^{-1} n^{-1}\Xi'\hat{\varepsilon} \\
&= \left(n^{-1} \begin{bmatrix} X' \\ E' \end{bmatrix} \begin{bmatrix} X & E \end{bmatrix} \right)^{-1} n^{-1} \begin{bmatrix} X' \\ E' \end{bmatrix} \hat{\varepsilon} \\
&= \begin{bmatrix} n^{-1}X'X & n^{-1}X'E \\ n^{-1}E'X & n^{-1}X'X \end{bmatrix}^{-1} \begin{bmatrix} n^{-1}X'\hat{\varepsilon} \\ n^{-1}E'\hat{\varepsilon} \end{bmatrix} \\
&= \hat{B}^{-1} \begin{bmatrix} n^{-1}X'\hat{\varepsilon} \\ n^{-1}E'\hat{\varepsilon} \end{bmatrix}
\end{aligned}$$

Now, notice:

$$\begin{aligned}
X'\hat{\varepsilon} &= X'(y - X\hat{\beta}) = X'y - X'X\hat{\beta} \\
&= X'y - X'y = 0
\end{aligned}$$

And:

$$E'\hat{\varepsilon} = n^{-1} \begin{bmatrix} n^{-1} \sum_{t=1}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_{t-1} \\ \vdots \\ n^{-1} \sum_{t=1}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_{t-p} \end{bmatrix} = \begin{bmatrix} \hat{\gamma}_1 \\ \vdots \\ \hat{\gamma}_p \end{bmatrix} = \hat{\gamma}$$

So:

$$\hat{\alpha} = \hat{B}^{-1} \begin{bmatrix} 0 \\ \hat{\gamma} \end{bmatrix}$$

(b): Note that convergence in probability of a partitioned matrix is equivalent to convergence in probability of the individual components. Hence, to show that $\hat{B} \xrightarrow{p} B$, we can show that the components of $\hat{B}$ converge in probability to the corresponding components of B.

First, notice that $n^{-1}X'X = n^{-1}\sum_{t=1}^{n} x_t x_t'$. Since $\{(y_t, x_t)\}$ is stationary ergodic, $\{x_t x_t'\}$ is stationary ergodic. Moreover, the mean exists, so by the WLLN for stationary ergodic processes:

$$n^{-1}X'X = n^{-1} \sum_{t=1}^{n} x_t x_t' \xrightarrow{p} \mathbb{E}[x_t x_t']$$

Next, notice that:

$$
\begin{aligned}
n^{-1}X'E &= n^{-1}\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}\begin{bmatrix} 0 & 0 & \cdots & 0 \\ \hat{\varepsilon}_1 & 0 & \cdots & 0 \\ \hat{\varepsilon}_2 & \hat{\varepsilon}_1 & \cdots & 0 \\ \hat{\varepsilon}_3 & \hat{\varepsilon}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\varepsilon}_{n-1} & \hat{\varepsilon}_{n-2} & \cdots & \hat{\varepsilon}_{n-p} \end{bmatrix} \\
&= \begin{bmatrix} n^{-1}\sum_{t=1}^{n} x_t\hat{\varepsilon}_{t-1} & \cdots & n^{-1}\sum_{t=1}^{n} x_t\hat{\varepsilon}_{t-p} \end{bmatrix}
\end{aligned}
$$

Now, for $k \in [p]$:

$$
\begin{aligned}
x_t\hat{\varepsilon}_{t-k} &= x_t(y_{t-k} - x_{t-k}'\hat{\beta}) \\
&= x_t(x_{t-k}'\beta + \varepsilon_{t-k} - x_{t-k}'\hat{\beta}) \\
&= x_t\varepsilon_{t-k} - x_t x_{t-k}'(\hat{\beta} - \beta)
\end{aligned}
$$

So:

$$
n^{-1}\sum_{t=1}^{n} x_t\hat{\varepsilon}_{t-k} = n^{-1}\sum_{t=1}^{n} x_t\varepsilon_{t-k} - \left(n^{-1}\sum_{t=1}^{n} x_t x_{t-k}'\right)(\hat{\beta} - \beta)
$$

Again, given ergodic stationarity and assuming the relevant moments exist:

$$
x_t\hat{\varepsilon}_{t-k} \xrightarrow{p} \mathbb{E}[x_t\varepsilon_{t-k}] - \mathbb{E}[x_t x_{t-k}'] \cdot 0 = \mathbb{E}[x_t\varepsilon_{t-k}]
$$

Hence:

$$
n^{-1}X'E \xrightarrow{p} \begin{bmatrix} \mathbb{E}[x_t\varepsilon_{t-1}] & \cdots & \mathbb{E}[x_t\varepsilon_{t-p}] \end{bmatrix}
$$

Finally, notice that:

$$
\begin{aligned}
n^{-1}E'E &= n^{-1}\begin{bmatrix} 0 & \hat{\varepsilon}_1 & \hat{\varepsilon}_2 & \hat{\varepsilon}_3 & \cdots & \hat{\varepsilon}_{n-1} \\ 0 & 0 & \hat{\varepsilon}_1 & \hat{\varepsilon}_2 & \cdots & \hat{\varepsilon}_{n-2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \hat{\varepsilon}_{n-p} \end{bmatrix}\begin{bmatrix} 0 & 0 & \cdots & 0 \\ \hat{\varepsilon}_1 & 0 & \cdots & 0 \\ \hat{\varepsilon}_2 & \hat{\varepsilon}_1 & \cdots & 0 \\ \hat{\varepsilon}_3 & \hat{\varepsilon}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\varepsilon}_{n-1} & \hat{\varepsilon}_{n-2} & \cdots & \hat{\varepsilon}_{n-p} \end{bmatrix} \\
&= \begin{bmatrix} n^{-1}\sum_{t=1}^{n}\hat{\varepsilon}_{t-1}^2 & \cdots & n^{-1}\sum_{t=1}^{n}\hat{\varepsilon}_{t-1}\hat{\varepsilon}_{t-p} \\ \vdots & \ddots & \vdots \\ n^{-1}\sum_{t=1}^{n}\hat{\varepsilon}_{t-p}\hat{\varepsilon}_{t-1} & \cdots & n^{-1}\sum_{t=1}^{n}\hat{\varepsilon}_{t-p}^2 \end{bmatrix}
\end{aligned}
$$

Now, for $j, k \in [n]$:

$$
\hat{\varepsilon}_{t-j}\hat{\varepsilon}_{t-k} = x_{t-j}'(\beta - \hat{\beta})x_{t-k}'(\beta - \hat{\beta}) + x_{t-j}'(\beta - \hat{\beta})\varepsilon_{t-k} + \varepsilon_{t-j}x_{t-k}'(\beta - \hat{\beta}) + \varepsilon_{t-j}\varepsilon_{t-k}
$$

Clearly, given $\hat{\beta} \xrightarrow{p} \beta$, we have:

$$n^{-1}\hat{\varepsilon}_{t-j}\hat{\varepsilon}_{t-k} \xrightarrow{p} \mathbb{E}[\varepsilon_{t-j}\varepsilon_{t-k}] = \begin{cases} \sigma^2 & j = k \\ 0 & j \neq k \end{cases}$$

So:

$$n^{-1}E'E \xrightarrow{p} \sigma^2 I$$

Thus, we have shown:

$$\hat{B} = \begin{bmatrix} n^{-1}X'X & n^{-1}X'E \\ n^{-1}E'X & n^{-1}E'E \end{bmatrix} \xrightarrow{p} \begin{bmatrix} \mathbb{E}[x_t x_t'] & H \\ H' & \sigma^2 I \end{bmatrix} = B$$

(c): Recall, $\hat{\alpha} = \hat{B}^{-1}\begin{bmatrix} 0 & \hat{\gamma} \end{bmatrix}'$. Since matrix inversion is continuous, by the continuous mapping theorem and the previous result, $\hat{B}^{-1} \xrightarrow{p} B^{-1}$. And, given that $\hat{\gamma}$ is consistent for $\gamma = 0$:

$$\hat{\alpha} \xrightarrow{p} B^{-1}\begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0$$

(d): Let $\hat{u}$ denote the vector of residuals in the auxiliary regression. Then:

$$\begin{aligned}
\text{SSR} &= \hat{u}'\hat{u} \\
&= (\hat{\varepsilon} - \Xi\hat{\alpha})'(\hat{\varepsilon} - \Xi\hat{\alpha}) \\
&= \hat{\varepsilon}'\hat{\varepsilon} - 2\hat{\alpha}\Xi'\hat{\varepsilon} + \hat{\alpha}\Xi'\Xi\hat{\alpha} \\
&= \hat{\varepsilon}'\hat{\varepsilon} - \hat{\alpha}(2\Xi'\hat{\varepsilon} - \Xi'\Xi\hat{\alpha}) \\
&= \hat{\varepsilon}'\hat{\varepsilon} - \hat{\alpha}\Big(2\Xi'\hat{\varepsilon} - \Xi'\Xi(\Xi'\Xi)^{-1}\Xi'\hat{\varepsilon}\Big) \\
&= \hat{\varepsilon}'\hat{\varepsilon} - \hat{\alpha}\Xi'\hat{\varepsilon} \\
&= \hat{\varepsilon}'\hat{\varepsilon} - \hat{\alpha}\begin{bmatrix} X & E \end{bmatrix}'\hat{\varepsilon} \\
&= \hat{\varepsilon}'\hat{\varepsilon} - \hat{\alpha}\begin{bmatrix} X'\hat{\varepsilon} \\ E'\hat{\varepsilon} \end{bmatrix} \\
&= \hat{\varepsilon}'\hat{\varepsilon} - \hat{\alpha}\begin{bmatrix} 0 \\ n\hat{\gamma} \end{bmatrix} \\
\implies n^{-1}\text{SSR} &= n^{-1}\sum_{t=1}^{n}\hat{\varepsilon}_t^2 - \hat{\alpha}\begin{bmatrix} 0 \\ \hat{\gamma} \end{bmatrix} \\
&\xrightarrow{p} \mathbb{E}[\hat{\varepsilon}_t^2] - 0 \cdot \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
&= \sigma^2
\end{aligned}$$

(e): In matrix form, the hypothesis that the coefficients on the lagged residuals are equal to zero is:

$$\begin{bmatrix} 0_{p\times K} & I_{p\times p} \end{bmatrix}\alpha = 0$$

Letting $R = \begin{bmatrix} 0 & I \end{bmatrix}$, the F statistic is defined as:

$$F = \frac{(R\hat{\alpha})'\Big[R(\Xi'\Xi)^{-1}R'\Big]^{-1}R\hat{\alpha}/p}{\text{SSR}/(n-K-p)}$$

Notice:

$$R\hat{\alpha} = \begin{bmatrix}0 & I\end{bmatrix} \hat{B}^{-1} \begin{bmatrix}0 \\ \hat{\gamma}\end{bmatrix} = R\hat{\alpha} = \begin{bmatrix}0 & I\end{bmatrix} \begin{bmatrix}\hat{B}^{11} & \hat{B}^{12} \\ \hat{B}^{21} & \hat{B}^{22}\end{bmatrix} \begin{bmatrix}0 \\ \hat{\gamma}\end{bmatrix} = \hat{B}^{12}\hat{\gamma}$$

And:

$$\begin{aligned} R(\Xi'\Xi)^{-1}R' &= \begin{bmatrix}0 & I\end{bmatrix} (n\hat{B})^{-1} \begin{bmatrix}0 \\ I\end{bmatrix} \\ &= n^{-1} \begin{bmatrix}0 & I\end{bmatrix} \begin{bmatrix}\hat{B}^{11} & \hat{B}^{12} \\ \hat{B}^{21} & \hat{B}^{22}\end{bmatrix} \begin{bmatrix}0 \\ I\end{bmatrix} \\ &= n^{-1}\hat{B}^{22} \end{aligned}$$

Hence:

$$\begin{aligned} F &= \frac{n\gamma'(\hat{B}^{22})^{-1}\hat{B}^{22}\hat{\gamma}/p}{\mathrm{SSR}/(n-K-p)} \\ \implies pF &= \frac{\hat{\gamma}'\hat{B}^{22}\hat{\gamma}}{\mathrm{SSR}/(n-K-p)} \end{aligned}$$

(f): This follows trivially from the formula for calculating the inverse of a partitioned matrix, which provides that the lower right block of the $\hat{B}^{-1}$, in terms of the components of $\hat{B}$, is:

$$\left(n^{-1}E'E - n^{-1}E'X(n^{-1}X'X)^{-1}n^{-1}X'E\right)^{-1}$$

(g): Consider the probability limit of $\hat{B}^{22} = \left[n^{-1}E'E - n^{-1}E'X(n^{-1}X'X)^{-1}n^{-1}X'E\right]^{-1}$. As we've shown, $n^{-1}E'E \xrightarrow{p} \sigma^2 I$ and $n^{-1}X'X \xrightarrow{p} \mathbb{E}[x_t x_t']$. Now, notice:

$$n^{-1}E'X = \begin{bmatrix} n^{-1}\sum_{t=1}^{n} x_t\hat{\varepsilon}_{t-1} \\ \vdots \\ n^{-1}\sum_{t=1}^{n} x_t\hat{\varepsilon}_{t-p} \end{bmatrix}$$

For $k \in [p]$, $x_t\hat{\varepsilon}_{t-k} = x_t x_{t-k}'(\beta - \hat{\beta}) + x_t\varepsilon_{t-k}$. Assuming $\mathbb{E}[x_t x_{t-k}']$ is finite for all $k \in [p]$, by standard arguments, we have:

$$n^{-1}\sum_{t=1}^{n} E'X \xrightarrow{p} \begin{bmatrix} \mathbb{E}[x_t\varepsilon_{t-1}] \\ \vdots \\ \mathbb{E}[x_t\varepsilon_{t-p}] \end{bmatrix}$$

It thus follows, by the continuous mapping theorem, that:

$$\hat{B}^{22} \xrightarrow{p} (I - \Phi)^{-1}/\sigma^2$$

where Φ is defined elementwise by:

$$\phi_{jk} = \mathbb{E}[x_t\varepsilon_{t-j}](\mathbb{E}[x_t x_t'])^{-1}\mathbb{E}[x_t\varepsilon_{t-k}]$$

Noting that $\text{SSR}/(n-K-p) = s^2 \xrightarrow{p} \sigma^2$ and $\hat{\gamma} \xrightarrow{p} \gamma$, we have that:

$$n^{-1}pF = \frac{\hat{\gamma}'\hat{B}^{22}\hat{\gamma}}{\text{SSR}/(n-K-p)} \xrightarrow{p} \hat{\gamma}'(I-\Phi)^{-1}\hat{\gamma}/\sigma^4$$

It is trivial to see that this is also the probability limit of Q^*_{BP}. Hence, $Q^*_{BP} - pF \xrightarrow{p} 0$. $\square$