Foundation of Econometrics - Homework

Giovanni Maffei, Timo Predöhl

November 20, 2022

Contents

1 Exercise 1: Spurious Regressions

1.1 Set-up and Definitions

- We are studying two independent stochastic processes: y_t and x_t with $t = 1, ..., T$.
- Each characterization of these processes is assumed to have independent errors: $u_{y,t} \overset{iid}{\sim} N(0,1)$ and $u_{x,t} \overset{iid}{\sim} N(0,1)$.

  Both errors are standardized random walk stochastic processes.

- The different regression modeles we are testing here result from the two processes below:

  $y_t = y_{t-1} + u_{y,t}$ and $x_t = x_{t-1} + u_{x,t}$

  x_t and y_t are integrated of order 1: I(1) and contain a unit root.

  $y_t = 0.8y_{t-1} + u_{y,t}$ and $x_t = 0.8x_{t-1} + u_{x,t}$.

  Both processes are represented as stationary AR(1) models[1] with an ergodic expected value of 0, as no intercept is modelled into those.

- Each regression model is estimated with OLS under asymptotic theory[2]:

  A.1: the combination of (y_t, k_t) for $t = 1, ..., T$ is IID, hence $(v_t, k_t) \sim IID$[3].

  A.2: regressors uncorrelated with the simultaneous error term, hence $E[v_t k_t] = 0_J$[4]. This assumption is not restricting the utilization of lagged values of the dependent variable into the regressors matrix.

  A.3: The VaR-Cov matrix of the regressors: $Q = E[k_t k_t']$ exists, is positive definite and has rank corresponding to the total number of regressors (no regressor is a linear combination of the others).

  A.4: $E[v_t^2 k_t k_t']$ exists and is positive definite.

[1] $\phi = 0.8 < \mid 1 \mid$

[2] Indeed, we are not assuming strict exogeneity of the regressors as we are using lagged values of the dependent variable as a regressor.

[3] By K we refer to the matrix of the regressors and v_t corresponds to the error term in each model.

[4] J represents the total number of regressors.

- The Monte Carlo exercise is based on $N_{sim} = 10,000$.
- We will test each regression model assuming different samples of pairs of observations: $t = 6, 12, 60, 120, 240, 360, 480$.

1.2 Part 1: Replication of the exercise for Figure 14.1 in Davidson MacKinnon (2005, book)

The four data generating processes (DPG) we are studying are:

- DPG_1: $y_t = \beta_1 + \beta_2 x_t + v_t$

  We assume that both x_t and y_t are I(1).

- DPG_2: $y_t = \beta_1 + \beta_2 x_t + v_t$

  We assume that both x_t and y_t are stationary AR(1) with $\phi = 0.8$.

- DPG_3: $y_t = \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + v_t$

  We assume that both x_t and y_t are I(1).

- DPG_4: $y_t = \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + v_t$

  We assume that both x_t and y_t are stationary AR(1) with $\phi = 0.8$.

In light of the hypothesis formulated in the previous section, we would expect that the estimates of β_2 and R^2 were not significantly different from 0[5]. Out of the Monte Carlo simulation performed by Davidson and MacKinnon (2005, book), we present three results[6]:

- Distribution of the R_2 for each sample size t
  Placeholder for chart 1!
- Distribution of the realizations of the t-statistics for the test of hypothesis: $H_0 : \beta_2 = 0$
  Placeholder for chart 2!
- Empirical rejection frequencies calculated as the ratio between the number of times H_0 got rejected against the variable t, assuming 5% as significance level. The size of the Monte Carlo test (N_{sim}) corresponds

[5] Such consideration for the R_2 does not apply directly to DPGs 3 and 4 where we are regressing on $L(y_t)$.

[6] The R code is available as an attachment to this submission.

to the total number of observations used to calculate the empirical rejection ratio for each point in the figure below. Hence, the higher this parameter and the lower the dispersion of the results in repeated samples of this simulation will be.
Placeholder for chart 3!

1.3 Part 2: Summarize the problems of spurious regressions in econometrics

The results obtained are different from what we previously expected. Indeed, by looking at Figure 1.3, we can observe that:

- Apart from DPG 4 whose H_0 rejection rate converges asymptotically to 5%, the empirical rejection rate is structurally higher than the significance level for the three other processes.
- Additionally, we notice that the proportion of rejections keeps increasing alongside the sample size for DPG 1 always implying a statistically significant relationship bwetween y_t and x_t even if this does not exist.
- Therefore, it emerges that the actual probability of a Type 1 error[7] is significantly higher than the assumed test size.

Usually, the bias of the β multipled by $n^{0.5}$ is $O_p(1)$, as per below demonstration and noting that $K'K = O_p(T)$ and $K'v = O_p(T^{0.5})$:

$$\begin{aligned}
\hat{\beta} - \beta &= (K'K)^{-1}K'v \\
T^{0.5}(\hat{\beta} - \beta) &= T^{0.5}((K'K)^{-1}K'v) \\
&= T^{0.5}O_p(T^{-1}O_p(T^{0.5}) \\
&= O_p(1)
\end{aligned}$$

The bias is eventually bounded for $T \to +\infty$, so it does not explode asymptotically.
However, we have here an issue of unit roots and spurious regressions. If, at least, one of the regressors is a unit root[8], we can derive that $E[K_t'K_t]$

[7] $P(R_{H_0}|H_0)$.

[8] This happens for DPG 1 and 3, although for the latter the proportion of rejections does not coverge to 1 as t increases. This is because we are reducing y_t by one lag turning the dependent variable into a white noise random walk.

is not a finite positive definite matrix anymore and there is a violation of the OLS assumptions under asymptotic theory (steps below are for DPG 1[9], same holds in case of multiple regressors):

$$
\begin{aligned}
E[x_T' x_T] &= \sum_{t=1}^{T} x_t^2 \\
&= T + (T-1) + ... + 2 + 1 \\
&= \frac{T(T+1)}{2}
\end{aligned}
$$

It follows that: i) $T^{-1}x_T'x_T$ is $O(T)$ and ii) this metric does not have a finite probability limit ($T \to +\infty$).
The distribution of the t-statistics does not converge to the Student's t even asymptotically causing an overrejection of the null.
Eventually, we observe that the issue of spurious regressions occurs even if all variables are stationary[10]. Although the proportion of rejections converges to a fixed number and $E[K_t'K_t]$ is a finite definite positive metric, we can see that the result for DPG 2 implies an overrejection of the null while the empirical frequency for DPG 4, the only correct regression model, converges to the 5% significance level only after increasing significantly the sample size (in the order of a few thousand observations).
The distortion for DPG 3 arises from the fact that neither the constant nor x_t have any explanatory power for y_t, therefore $y_t = v_t = 0.8y_{t-1} + u_{y,t}$.
As y_t is modelled as an AR(1), the error term of DPG 3 becomes:

$$
\begin{aligned}
v_t &= 0.8y_{t-1} + u_{y,t} \\
v_t &= 0.8(u_{y,t-1} + 0.8y_{t-2}) + +u_{y,t} \; t = 1, ..., T
\end{aligned}
$$

This intuition suggests that there is serial autocorrelation in the errors and the (asymptotic) solution to avoid too small standard errors in finite samples is to switch from the OLS VaR-Cov estimator to the Newey-West heteroskedasticity and autocorrelation consistent (HAC) one.

[9] $V(x_t) = E(x_t^2) = t$.
[10] AR(1) in this case.

2 Exercise 2: Univariate Time Series Regression and Jensen's alpha

2.1 Set-up and Definitions

- We are assuming a timeseries regression for one asset: $i = 1, ..., N$, with N corresponding to the total number of securities available in the investable universe:

$$r_{it} = \alpha_i + \beta_{im} z_{mt} + u_{it} \quad t = 1, ..., T \tag{1}$$
$$u_i | z_{mt} \overset{iid}{\sim} (0, w_i^2)$$

- X is the matrix of the regressors.
- $E[x_i x_i']$ exists, and is definite positive.
- $Rank(X) = 2$ (no redunant regressors), hence $(X'X)^{-1}$ exists.
- $\hat{\alpha}_i$ and $\hat{\beta}_{im}$ are the OLS estimators of α and β in 1.
- z_{mt} corresponds to the return of a market index (used as proxy of the market portfolio) at time t.

2.2 Part 1: Demonstrate that $\hat{\alpha}_i = \overline{r}_i - \hat{\beta}_i \overline{z}_m$ and $\hat{\beta}_i = \frac{\hat{\sigma}_{im}}{\hat{\sigma}_m^2}$

Noting that if we substitute the intercept by adding 1_T as the first column of the regressors matrix: X, we can rewrite 1 in the following matricial representation: $r_{i,(1\times T)} = X_{i,(T\times 2)} \beta_{i,(2\times 1)} + u_{i,(T\times 1)}$, with X defined as a (1 x 2) matrix: $[1_{(Tx1)}, z_{m,(Tx1)}]$[11].
Additionally, recalling that: i) the OLS estimator of α_i and β_{im} corresponds to: $(X'X)^{-1}X'r_i$, ii) $1_T' 1_T = T$ and iii) the inverse of $(X'X)^{-1}$ is:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$
$$(X'X)^{-1} = \frac{1}{T\sum_{t=1}^T z_{mt}^2 - (\sum_{t=1}^T z_{mt})^2} \begin{bmatrix} \sum_{t=1}^T z_{mt}^2 & -\sum_{t=1}^T z_{mt} \\ -\sum_{t=1}^T z_{mt} & T \end{bmatrix} \tag{2}$$

[11]The second term of each matrix subscript is meant to represent the dimension of the corresponding metric for the avoidance of doubt, we will drop such notation for tractability reasons.

And observing $X'r_i = \begin{bmatrix} \sum_{t=1}^{T} r_{it} \\ \sum_{t=1}^{T} z_{mt} r_{it} \end{bmatrix}$, we derive that:

$$\hat{\beta}_i = (X'X)^{-1}X'r_i$$
$$= \begin{bmatrix} \frac{T(\frac{1}{T}\sum_{t=1}^{T} z_{mt}^2)T(\frac{1}{T}\sum_{t=1}^{T} r_{it}) - T(\frac{1}{T}\sum_{t=1}^{T} z_{mt})T(\frac{1}{T}\sum_{t=1}^{T} z_{mt}r_{it})}{T^2(\frac{1}{T}\sum_{t=1}^{T} z_{mt}^2) - T^2(\frac{1}{T}\sum_{t=1}^{T} z_{mt})^2} \\ \frac{T^2(\frac{1}{T}\sum_{t=1}^{T} z_{mt}r_{it}) - T(\frac{1}{T}\sum_{t=1}^{T} z_{mt})T(\frac{1}{T}\sum_{t=1}^{T} r_{it})}{T^2(\frac{1}{T}\sum_{t=1}^{T} z_{mt}^2) - T^2(\frac{1}{T}\sum_{t=1}^{T} z_{mt})^2} \end{bmatrix} \tag{3}$$

Besides, we can further refine this expression by noting:

- $\frac{1}{T}\sum_{t=1}^{T} z_{mt} = \overline{z}_m$
- $\frac{1}{T}\sum_{t=1}^{T} r_{it} = \overline{r}_i$
- $T^2(\frac{1}{T}\sum_{t=1}^{T} z_{mt}^2) - T^2(\frac{1}{T}\sum_{t=1}^{T} z_{mt})^2 = T^2\hat{\sigma}_m^2$, with $\hat{\sigma}_m^2$ as the variance of the market index returns.
- $T^2(\frac{1}{T}\sum_{t=1}^{T} z_{mt}r_{it}) - T(\frac{1}{T}\sum_{t=1}^{T} z_{mt})T(\frac{1}{T}\sum_{t=1}^{T} r_{it}) = T^2\hat{\sigma}_{z,i}$, with $\hat{\sigma}_{z,i}$ as the covariance between the returns of the single asset and the market index.
- $T(\frac{1}{T}\sum_{t=1}^{T} z_{mt}^2)T(\frac{1}{T}\sum_{t=1}^{T} r_{it}) - T(\frac{1}{T}\sum_{t=1}^{T} z_{mt})T(\frac{1}{T}\sum_{t=1}^{T} z_{mt}r_{it}) =$
  $= T^2\overline{r}_i(\frac{1}{T}(\sum_{t=1}^{T} z_{mt}^2 - \overline{z}_m{}^2) - T^2\overline{z}_m(\frac{1}{T}(\sum_{t=1}^{T} z_{mt}r_{it} - \overline{z}_m\overline{r}_i)$
  $= T^2(\overline{r}_i\hat{\sigma}_m^2 - \overline{z}_m\hat{\sigma}_{z,i})$

By substituting the formulae above into 3, we derive:

$$\hat{\beta}_i = \begin{bmatrix} \hat{\alpha}_i \\ \hat{\beta}_{im} \end{bmatrix}$$
$$= \begin{bmatrix} \frac{\overline{r}_i\hat{\sigma}_m^2 - \overline{z}_m\hat{\sigma}_{z,i}}{\hat{\sigma}_m^2} \\ \frac{\hat{\sigma}_{z,i}}{\hat{\sigma}_m^2} \end{bmatrix}$$
$$= \begin{bmatrix} \overline{r}_i - \overline{z}_m\hat{\beta}_{im} \\ \frac{\hat{\sigma}_{z,i}}{\hat{\sigma}_m^2} \end{bmatrix} \tag{4}$$

2.3 Part 2: Demonstrate the asymptotic distribution of $\hat{\beta}_i|z_m$

As we are not assuming any specific form for $u_i|z_{mt}$, we will study the distribution of $\hat{\beta}_i|z_m$ asymptotically ($T \to \infty$). Therefore, by rearranging the definition of $\hat{\beta}_i$ we derive:

$$\begin{aligned}\hat{\beta}_i|z_{mt} &= (X'X)^{-1}X'r_i = \beta_i + (X'X)^{-1}X'u_i \\ \sqrt{T}(\hat{\beta}_i - \beta_i) &= \sqrt{T}(X'X)^{-1}X'u_i \\ &= (\frac{X'X}{T})^{-1}\sqrt{T}X'u_i\end{aligned}$$

Further noting that (under $T \to +\infty$):

- $(\frac{X'X}{T}) \xrightarrow{p} E[x_i x_i']$ due to the application of the Law of Large Numbers (LLN) because $[x_{it}]_{t=1}^{T}$ is a sequence of IID observations with finite expected value and variance. Additionally, by using the Slutsky's theorem we obtain: $(\frac{X'X}{T})^{-1} \xrightarrow{p} E[x_i x_i']^{-1}$.
- $X'u_i \xrightarrow{d} X'N(0, w_i^2 I_T)X$ due to the application of the Central Limit Theorem (CLT) in light of: i) IID observations, ii) errors uncorrelated with the regressors and iii) finite moments[12].

By applying the properties of the multivariate Normal distribution, we obtain:

$$\hat{\beta}_i \xrightarrow{d} N(\beta_i, (X'X)^{-1}X'w_i^2 I_T X(X'X)^{-1})$$

Simplifying the variance term of the distribution of $\hat{\beta}_i$:

$$\begin{aligned}V(\hat{\beta}_i) &= (X'X)^{-1}X'w_i^2 I_T X(X'X)^{-1} \\ &= w_i^2(X'X)^{-1}\end{aligned}$$

Additionally, we can rewrite 2 as per below:

$$\begin{aligned}(X'X)^{-1} &= \frac{1}{T^2\hat{\sigma}_m^2}\begin{bmatrix}(\sum_{t=1}^{T} z_{mt}^2) - T\overline{z}_m^2 + T\overline{z}_m^2 & -T\overline{z}_m \\ -T\overline{z}_m & T\end{bmatrix} \\ &= \frac{1}{T^2\hat{\sigma}_m^2}\begin{bmatrix}(T\hat{\sigma}_m^2 + T\overline{z}_m^2) & -T\overline{z}_m \\ -T\overline{z}_m & T\end{bmatrix} \\ &= \frac{1}{T}\begin{bmatrix}(1 + \frac{\overline{z}_m^2}{\hat{\sigma}_m^2}) & \frac{-\overline{z}_m}{\hat{\sigma}_m^2} \\ \frac{-\overline{z}_m}{\hat{\sigma}_m^2} & \frac{1}{\hat{\sigma}_m^2}\end{bmatrix}\end{aligned} \tag{5}$$

[12] As mentioned in the *Set-up and Definitions* part of this exercise.

Eventually, we conclude this demonstration by expanding $\hat{\beta}_i$ into its components and by replacing 5 into $\hat{\beta}_i \xrightarrow{d} N(\beta_i, V(\hat{\beta}_i)$:

$$\begin{bmatrix} \hat{\alpha}_i \\ \hat{\beta}_{im} \end{bmatrix} \sim N(\begin{bmatrix} \alpha_i \\ \beta_{im} \end{bmatrix}, \frac{w_i^2}{T} \begin{bmatrix} (1 + \frac{\bar{z}_m^2}{\hat{\sigma}_m^2}) & \frac{-\bar{z}_m}{\hat{\sigma}_m^2} \\ \frac{-\bar{z}_m}{\hat{\sigma}_m^2} & \frac{1}{\hat{\sigma}_m^2} \end{bmatrix}) \tag{6}$$

2.4 Part 3: Demonstrate the t-statistics to test H_0: $\alpha_i = 0$

We add below a few considerations additional to what presented in *Set-up and Definitions* to support this proof:

- We are studying the asymptotic distribution of this test under $T \rightarrow +\infty$[13].
- This is a linear hypothesis test, with $m = 1$ linear restrictions tested[14].
- The generic matrix R (m x K), formalizing the linear restrictions on β, corresponds to the vector $\begin{bmatrix} 1 & 0 \end{bmatrix}$.
- The generic vector q (m x 1) represents the restricted values imposed on β. As we are imposing only one restriction, q corresponds to a scalar metric: 0.
- As we don't know the actual variance of $u_i|z_{mt}$ but we estimate the corresponding OLS residuals, we utilize $\frac{1}{T}\sum_{t=1}^{T} \hat{u}_{it}^2$ as an asymptotically unbiased estimator by applying the LLN[15].

Under the null hypothesis H_0: $R\beta_i = q$:

$$R\hat{\beta}_i - q = R(\hat{\beta}_i - \beta_i)$$

Furthermore, as proved in the previous exercise and by applying the properties of the multivariate Gaussian distribution, we derive:

$$R(\hat{\beta}_i - \beta_i) \sim N(0, RV(\beta_i)R')$$

$$[RV(\beta_i)R']^{-\frac{1}{2}} R(\hat{\beta}_i - \beta_i) \xrightarrow{d} N(0, I_m) \tag{7}$$

[13] Indeed, no distribution is provided for the errors in 1.

[14] The only restriction we are imposing on β_i is: $\alpha_i = 0$.

[15] $\overline{u}_i = 0$ because of the inclusion of the intercept in X.

By substituting for $R = \begin{bmatrix} 1 & 0 \end{bmatrix}$ in $RV(\beta_i)R'$ and rearranging the terms:

$$\begin{aligned} RV(\beta_i)R' &= \frac{{w_i}^2}{T} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} (1+\frac{\overline{z}_m^2}{\hat{\sigma}_m^2}) & \frac{-\overline{z}_m}{\hat{\sigma}_m^2} \\ \frac{-\overline{z}_m}{\hat{\sigma}_m^2} & \frac{1}{\hat{\sigma}_m^2} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ &= \frac{{w_i}^2}{T}(1+\frac{\overline{z}_m^2}{\hat{\sigma}_m^2}) \end{aligned}$$

Therefore 7 becomes:

$$\begin{aligned} [RV(\beta_i)R']^{-\frac{1}{2}} R(\hat{\beta}_i - \beta_i) &= \frac{R(\hat{\beta}_i - \beta_i)}{w_i\sqrt{\frac{(1+\frac{\overline{z}_m^2}{\hat{\sigma}_m^2})}{T}}} \\ &= \frac{\hat{\alpha}_i - \alpha|H_0}{w_i\sqrt{\frac{(1+\frac{\overline{z}_m^2}{\hat{\sigma}_m^2})}{T}}} \\ &= \frac{\hat{\alpha}_i}{w_i\sqrt{\frac{(1+\frac{\overline{z}_m^2}{\hat{\sigma}_m^2})}{T}}} \end{aligned}$$

Eventually by using the unbiased estimator for w_i^2, as mentioned at the beginning of this proof, we derive[16]:

$$\frac{\hat{\alpha}_i}{\hat{w}_i\sqrt{\frac{(1+\frac{\overline{z}_m^2}{\hat{\sigma}_m^2})}{T}}} \xrightarrow{d} N(0,1) \tag{8}$$

3 Exercise 3: Pricing errors in FM

3.1 Set-up and Definitions

- The data generating process (DGP) for the returns in excess of the risk-free rate with κ tradable factors is:

$$r_t = \beta f_t + u_t \quad t = 1, ..., T \tag{9}$$

- r_t is the (N x 1) vector containing the excess return of each security at time t: $r_{i,t}$ for i = 1,...,N, with N corresponding to the total number of securities available in the investable universe.

[16] I_m is the scalar 1, hence $N(0, I_m) = N(0, 1)$.

- u_t is the (N x 1) vector containing the errors of the DGP and is assumed: $\overset{iid}{\sim}$ $(0, \Omega)$ over t with finite matrix Ω.

- β is the (N x K)[17] matrix reporting on each line the factor loadings for i = 1,...,N[18].

  Rank(β) = K therefore, the inverse of $\beta'\beta$ exists and is definite positive.

- f_t is the (K x 1) vector of the excess returns of the tradable risk factors in t, distributed as per $\overset{iid}{\sim}$ (λ, Ω_f) over t. λ represents the (K x 1) vector of the true risk premia: $E(f_t) = \lambda$.[19]

- u_s is independent from f_t $\forall(s,t)$ hence, $E[u_s f_t] = 0$ (hypothesis of strict exogeneity of the regressors).

- We assume that the true factor loadings are known, hence we dont face a problem of error in variables (EIV) for this DPG[20].

- $\hat{\lambda}$ is estimated via OLS and corresponds to: $(\beta'\beta)^{-1}\beta' r_t$.

3.2 Part 1

Demonstrate that $\hat{\epsilon} \xrightarrow{p} 0$ as $T \to +\infty$ Given the OLS formula for the pricing errors of 9 is $\hat{\epsilon}_t = r_t - \beta\hat{\lambda}_t$ and replacing $\hat{\lambda}$ with the corresponding OLS result, we find that $\hat{\epsilon}_t = r_t - \beta(\beta'\beta)^{-1}\beta' r_t = M_\beta r_t$[21].
Therefore we observe that:

$$\hat{\epsilon} = \frac{1}{T}\sum_{t=1}^{T}\epsilon_t = \frac{1}{T}\sum_{t=1}^{T} M_b r_t \tag{10}$$

[17]N is deemed to be larger than K.

[18]We are including the vector of ones: $i_N = [1, 1, ..., 1]'$ as the first column of β in order to add the intercept to the DGP introduced above.

[19]As we are including the intercept in the DGP and r_t is the vector of the excess returns, we would expect $\lambda_{1,t}$ to be equal to 0.

[20]If β had not been known, we would have estimated the vector from the N timeseries regressions covering the entire sample of lenght: T.

[21]M_β corresponds to $I_N - \beta(\beta'\beta)^{-1}\beta$ which is symmetric and idempotent (I_N is the identity matrix (N x N)).

Looking at 10 and replacing r_t with 9, we derive:

$$\begin{aligned}\hat{\epsilon} &= \frac{1}{T}\sum_{t=1}^{T} M_b r_t \\ &= \frac{1}{T}\sum_{t=1}^{T} M_b(\beta f_t + u_t) \\ &= \frac{1}{T}\sum_{t=1}^{T} M_b(\beta f_t) + \frac{1}{T}\sum_{t=1}^{T} M_b u_t \end{aligned} \tag{11}$$

The first component of 11 is equal to 0 as it becomes: $\frac{1}{T}\sum_{t=1}^{T}(\beta f_t - \beta f_t)$, while for the second term we see:

$$\begin{aligned}\hat{\epsilon} &= \frac{1}{T}\sum_{t=1}^{T} M_b u_t \\ &= M_b(\frac{1}{T}\sum_{t=1}^{T} u_t)\end{aligned}$$

As $u_t \overset{iid}{\sim} (0, \Omega)$ and Ω is finite, we can apply the Law of Large Numbers (LLN) and obtain that $(\frac{1}{T}\sum_{t=1}^{T} u_t) \xrightarrow{p} 0$, it follows that: $\hat{\epsilon} \xrightarrow{p} 0$.

3.3 Part 2

Demonstrate that $\frac{1}{T}\sum_{t=1}^{T}(\hat{\epsilon}_t - \hat{\epsilon})(\hat{\epsilon}_t - \hat{\epsilon})' \xrightarrow{p} M_\beta \Omega M_\beta$ as $T \rightarrow +\infty$ Following the same setup and results of 1.2 and recalling that $(M_\beta)' = M_\beta$, we derive that:

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}(\hat{\epsilon}_t - \hat{\epsilon})(\hat{\epsilon}_t - \hat{\epsilon})' &= \frac{1}{T}\sum_{t=1}^{T}(M_\beta r_t)(M_\beta r_t)' \\ &= \frac{1}{T}\sum_{t=1}^{T} M_\beta(\beta f_t + u_t)(\beta f_t + u_t)' M_\beta \\ &= \frac{1}{T}\sum_{t=1}^{T} M_\beta(\beta f_t f_t' \beta' + u_t u_t' + 2\beta f_t u_t') M_\beta\end{aligned} \tag{12}$$

Additionally, we can decompose 12 into the following components:

- $\frac{1}{T}\sum_{t=1}^{T} M_\beta(\beta f_t f_t'\beta')M_\beta = \frac{1}{T}\sum_{t=1}^{T}(\beta f_t f_t'\beta' - \beta f_t f_t'\beta')M_\beta = 0$
- $\frac{1}{T}\sum_{t=1}^{T} M_\beta(2\beta f_t u_t')M_\beta = \frac{2}{T}\sum_{t=1}^{T}(\beta f_t u_t' - \beta f_t u_t')M_\beta = 0$
- $\frac{1}{T}\sum_{t=1}^{T} M_\beta u_t u_t' M_\beta$

Assuming: i) the fourth moment of u_t exists and is finite and ii) recalling that $u_t \overset{iid}{\sim} (0, \Omega)$, we can apply the LLN on $\frac{1}{T}\sum_{t=1}^{T} u_t u_t'$ and observe that this quantity $\xrightarrow{p} \Omega$. It follows that the last member of 12 $\xrightarrow{p} M_\beta \Omega M_\beta$, therefore:

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{\epsilon}_t - \hat{\epsilon})(\hat{\epsilon}_t - \hat{\epsilon})' \xrightarrow{p} M_\beta \Omega M_\beta \tag{13}$$