Advanced Econometrics
Assignment 1

Tibo Lachaert
Nicolás Romero Díaz

June 26, 2024

1   OLS Estimation

- *Replicate OLS results for the following cigarette demand model*

$$\ln\left(C_{it}\right) = \alpha + \rho\ln\left(C_{i,t-1}\right) + \beta_1\ln\left(P_{it}\right) + \beta_2\ln\left(Pn_{it}\right) + \beta_3\ln\left(Y_{it}\right) + \mu_{it} \tag{1}$$

- File: `OwnFunctions.R`

  - Line 4 in this file has the code for the `OLS_own` function, which manually implements OLS estimation.

- Implementation: `Case assignment - Dynamic panel data.R`

  - We present the results of our own implementation of the OLS estimator. Table 1 shows the results from running the code in this file until line 45.

	Coefs	Std. Dev	t-stats	p-values
$\ln\left(C_{i,t-1}\right)$	0.969	0.006	157.669	0.000
$\ln\left(P_{it}\right)$	-0.090	0.015	-6.183	0.000
$\ln\left(Pn_{it}\right)$	0.024	0.013	1.827	0.068
$\ln\left(Y_{it}\right)$	-0.031	0.006	-5.089	0.000

Table 1: OLS results from applying the `OLS_own` function.

2   Fixed Effects estimator

- *Write a Matlab/R function that implements the FE estimator*

- *Replicate the FE results reported by Baltagi*

2.1   Implementing the FE estimator

- File: `OwnFunctions.R`

  - Line 39 in this file has the code for the implementation of the Fixed Effects (FE) estimator. The function is called `FE_own`

- Implementation: `Case assignment - Dynamic panel data.R`

  - Running the code in the file above up to line 68, we obtain the estimates presented in Table 2

	Coefs	Std. Dev	t-stats	p-values
$\ln\left(C_{i,t-1}\right)$	0.833	0.013	66.269	0.000
$\ln\left(P_{it}\right)$	-0.299	0.024	-12.638	0.000
$\ln\left(Pn_{it}\right)$	0.034	0.027	1.238	0.216
$\ln\left(Y_{it}\right)$	0.100	0.024	4.200	0.000

Table 2: FE results from applying the `FE_own` function.

2.2 Simulating the properties of the FE estimator

- *Set up a Monte Carlo experiment to numerically illustrate the properties of the FE estimator in a dynamic setting*

$$y_{it} = \rho y_{i,t-1} + \varepsilon_{it} \tag{2}$$

$$\varepsilon \sim \mathcal{N}\left(0,1\right)$$

  - *Use a burn-in of 25 periods*
  - *Use different values of $\rho = 0, 0.5, 0.9$*
  - *Use all combinations of $T = 4, 10, 20, 50$ and $N = 10, 100$*

- Implementation: `Case assignment – Dynamic panel data.R`

  - Continuing to execute the code in this file until line 168 will implement a Monte Carlo simulation with the respective number of individuals (N) and time periods (T) for each of the ρ parameters.

  - Figure 1 summarizes the results obtained. Each subplot corresponds to a different analytical value of ρ. For a better comparison, the y-axis shows the estimated values $\hat{\rho}$ centered around the true value, i.e. $\hat{\rho} - \rho$. This value is represented by a red line in every subplot. The blue circle indicates the results for the panel with $N = 10$, while the orange crosses indicate results for $N = 100$.

  - For a given T, Figure 1 does not show any obvious difference in the estimations of $\hat{\rho}$ for values of $N = 10, 100$. On the other hand, we can see that increasing values of T result in closer estimations of the $\hat{\rho}$ parameter. This result was expected due to the analytical form of the Nickell bias, which is inversely proportional to T. As a result, the bias decreases with an increasing T. This bias is also shown graphically in Figure 1, since estimations for all T, N considered lie below the analytical value.
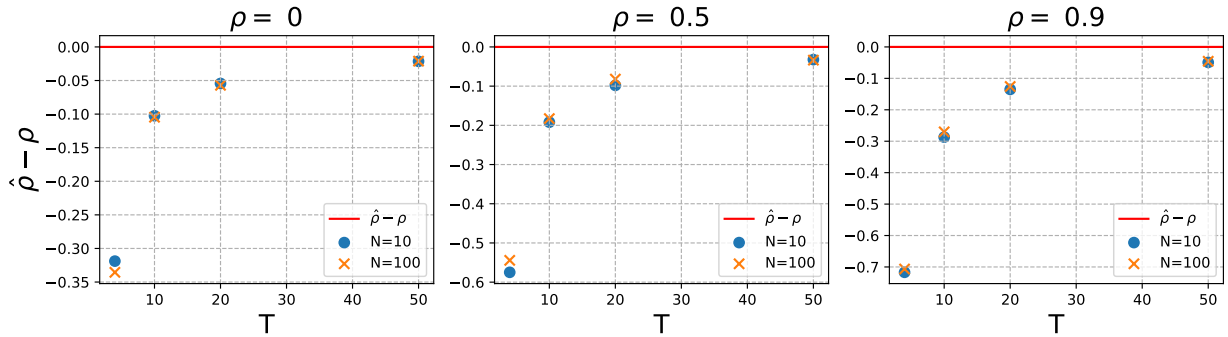

Figure 1: Results of Monte Carlo simulations of the FE estimator. Each subplot corresponds to different analytical values of ρ. The red line indicates the true value of the parameter. Blue dots indicate results for $N = 10$, while orange crosses present results for $N = 100$.

- *Compare the mean of the estimated standard error $\hat{se}(\hat{\rho}_{FE})$ to the true standard error $se(\hat{\rho}_{FE})$*
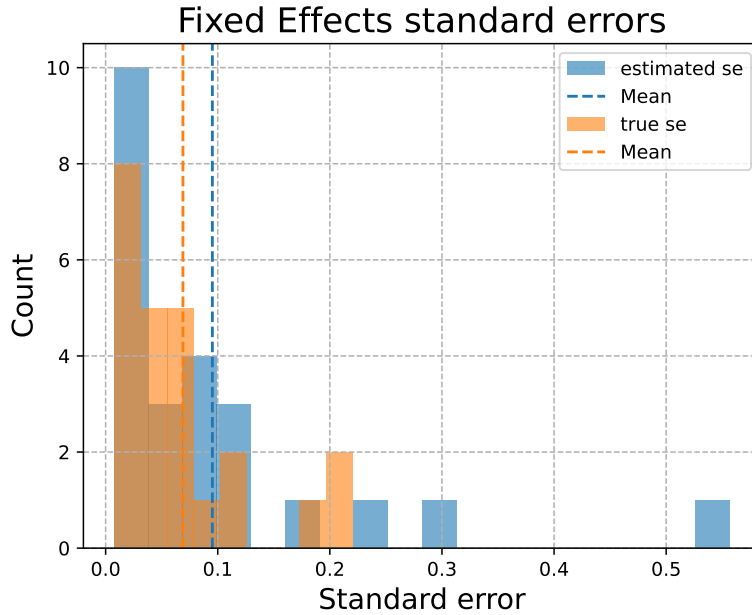
2

Figure 2: Blue: estimated standard errors; Orange: True standard errors. The dashed blue and orange lines correspond to the means of the estimated standard errors and the true standard errros, respectively.

- Files: `Case assignment - Results.xlsx` – Jupyter notebook: `plots.ipynb`

  - From the results presented in the files above, we calculates the means of the estimated and true standard errors to be 0.0952 and 0.0690. The average of the estimated standard errors is larger by 0.0262, which is a difference of 38.05%.
  - Figure 2 shows the results for the standard errors in our implementation. The distribution of these errors is relatively similar, and it appears that some outliers in the estimated se are the ones that cause the average of estimated standard errors to be larger than the true standard errors.

2.3   Simulation methods for bias-correction and inference

In this section, we implement a half-panel jackknife to bias-correct a FE estimator.

- *Implement a half-panel jackknife bias-corrected FE estimator*

- *Use the Monte Carlo setup to compare the mean of $\hat{\rho}_{FE}$ to the population parameter ρ*

- Implementation: `jackknife_bias_correction.R`

  - The code in this file starts with the function `generate_ar1_panel`, which generates an AR1 panel dataset as described by equation (2).
  - In line 42, the function `create_half_panel` splits a panel into two halves along the time dimension, under the condition that the total time T is even.
  - Line 118 defines the `block_bootstrap` function, which ensures that the resampling procedure in the bootstrap does not interfere with the persistence pattern ρ in the AR1 process.
  - Line 158 applies the jackknife procedure for the values of ρ, T, and N given in question 2.2
  - Figure 3 presents the results for the bias-corrected estimator. This figure reproduces the presentation of results for question 2.2, yet, in this case, there is no apparent pattern in the estimated $\hat{\rho}$ as N and T change. We can see that in several occasions, the jackknife procedure results in an over-correction of the estimated persistence parameter.
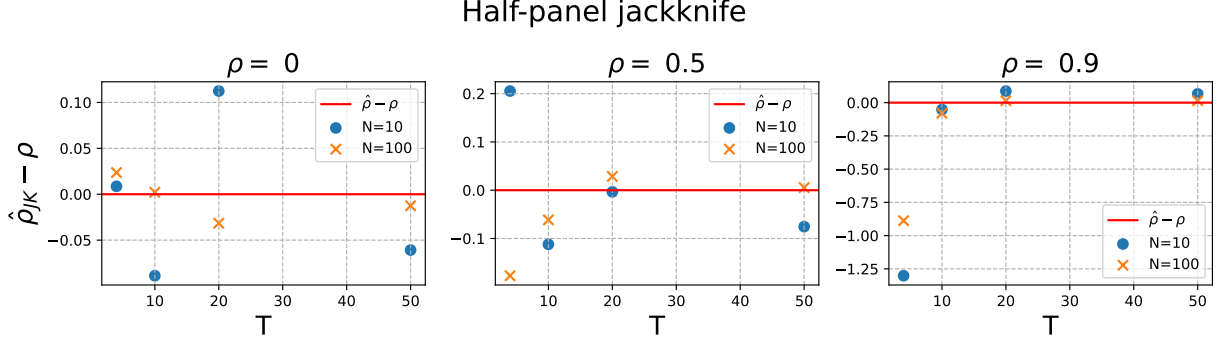
3

Figure 3: Results of the half-panel jackknife bias correction. Each subplot corresponds to different analytical values of ρ. The red line indicates the true value of the parameter. Blue dots indicate results for $N = 10$, while orange crosses present results for $N = 100$.

- To better compare between the non-corrected FE estimator and the bias-corrected version, we present a histogram of the difference between the estimated values and the true parameter ($\hat{\rho} - \rho$). The distribution of this error is presented in Figure 4. We can see that the distribution of the bias-corrected estimator is less skewed than for the vanilla FE. Other than 2 outliers, the jackknife procedure generates a relatively symmetric distribution around the analytical value of ρ.

- *Compare the mean of the bootstrapped standard error $se^b\left(\hat{\rho}_{FE_j}\right)$ to the true standard error $se\left(\hat{\rho}_{FE_j}\right)$*

- Implementation: `jackknife_bias_correction.R`

  - In our case, the mean of the bootstrapped se and the true se are practically equal, with each value being $se^b\left(\hat{\rho}_{FE_j}\right) = 0.0577$ and $se\left(\hat{\rho}_{FE_j}\right) = 0.0579$. These results are calculated in lines 213 and 216 of the file denoted above.
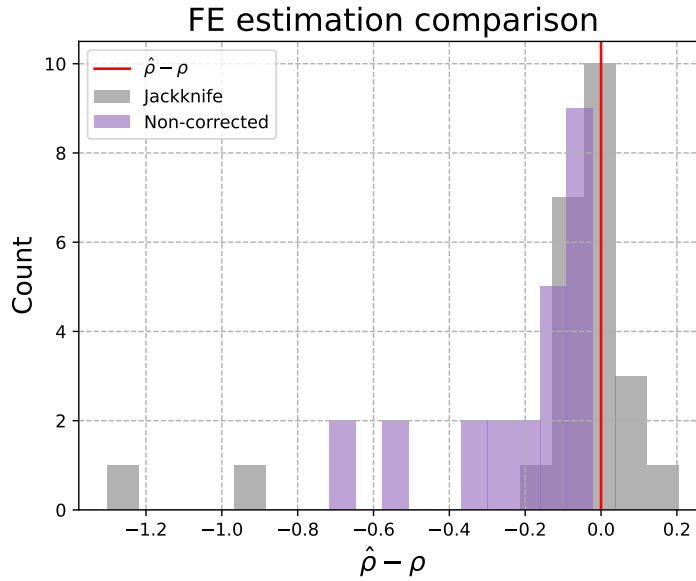

Figure 4: Distribution of the estimated $\hat{\rho}$ for jackknife (grey) and non-corrected FE (purple). The distribution is centered around the analytical value of ρ, which translates all estimated values around zero.

3 Implementing the panel GMM estimator

- *Write an R function that implements the GMM estimator. Results should include one-step GMM estimates, t-stats, p-values. For overidentified models: Two-step GMM estimates, t-stats, p-values, Sargan/Hansen test*

- File: `OwnFunctions.R`

  - In line 95 of this file, the `GMM_own` function applies the GMM estimator as a one-step and/or two-step estimator and applies the Sargan-Hansen test.

- *Take first differences to obtain $\Delta y_{it} = \rho \Delta y_{i,t-1} + \Delta \varepsilon_{it}$*

- *Set up two different GMM estimators by using $\Delta y_{i,t-2}$ and $\Delta y_{i,t-2}, \Delta y_{i,t-3}$ as instruments for $\Delta y_{i,t-1}$*

- *Theoretically argue whether/when these instruments are strong or weak, valid*

- Implementation: `Case assignment - Dynamic panel data.R`

  - Lines 247 and 311 of this file set up the GMM estimation for the models with one and two instruments, respectively. Running this file until the end will produce an Excel file with results for all models analyzed. This file is the `Case assignment - Results.xlsx` file.

  - Using lagged values $y_{i,t-1}$, $y_{i,t-2}$ can be justified when we take that the error term is serially uncorrelated. When this is the case, the correlation between the lagged values $y_{i,t-1}$, $y_{i,t-2}$ and Δy_{it} depends on the value of the ρ parameter in the AR1 process. When this value of ρ is high, the instruments are stronger, and vice-versa.

- *Compare the mean of $\hat{\rho}_{GMM}$ to the population parameter ρ*

  - Figures 5 and 6 present the difference between the estimations of $\hat{\rho}_{GMM}$ for the one-step and two-step estimators respectively. The top part of each figure presents the full results, while the bottom part removes the largest outliers for $N = 10, 100$. The GMM estimator seems to be particularly sensitive to the N parameter. We can see this in each subplot, where the larger $N = 100$ value produces an estimate that is much closer to the true value ρ in comparison to the smaller, $N = 10$ case. This observation seems to be confirmed by the dashed lines, which indicate the average values of $\hat{\rho}$, with the orange line ($N = 100$) being closer to the red line than the blue line ($N = 10$) in every subplot. This pattern holds true both for the one-step and two-step estimators presented in figures 5, 6.

- *Compare the true standard deviation to that of the FE estimator*

  - Figure 7 presents a histogram of the one-step and two-step GMM estimator in blue and green, respectively. These are the standard errors for all values of N, T, and ρ that we have calculated. The left subplot shows the distribution for the one-step GMM, while the right subplot presented the two-step GMM. The mean of these standard errors, calculated near the end of the jupyter notebook "`plots.ipynb`", are 0.1192 and 0.2761 for the one-step and two-step GMM, respectively. Meanwhile, the average se for Fixed Effects was 0.0690.

  - The accuracy of the two-step GMM was consistently better, as seen in Figures 5, 6. Nevertheless, the standard error of the two-step approach was much larger than the one-step GMM. The differences between the standard errors are $se_{GMM_1} - se_{FE} = 0.0502$ and $se_{GMM_1} - se_{FE} = 0.2071$, indicating that both GMM approaches produce larger standard errors than the FE estimator.

- *Compare the mean of the estimated standard error $\hat{se}(\hat{\rho}_{GMM})$ to the true standard error $se(\hat{\rho}_{GMM})$*

  - We take the average of the estimated standard errors for all values of ρ, T, N considered for the one-step and two-step GMM estimators. These averages are presented in Table 3, with the top section of the table presenting the average of all standard errors. Because of the large discrepancy in the values, the middle part of Table 3 shows these averages excluding the largest outlier. Even

	One-step GMM	Two-step GMM
$\hat{se}\left(\hat{\rho}_{GMM}\right)$	306.0061	0.3272
$se\left(\hat{\rho}_{GMM}\right)$	4.2480	0.2761
Difference	301.7581	0.0511
$\hat{se}\left(\hat{\rho}_{GMM}\right)$	12.3816	0.1863
$se\left(\hat{\rho}_{GMM}\right)$	0.9593	0.1995
Difference	11.4223	-0.0131
$\hat{se}\left(\hat{\rho}_{GMM}\right)$	0.1649	0.1398
$se\left(\hat{\rho}_{GMM}\right)$	0.1867	0.1617
Difference	-0.0217	-0.0219

Table 3: TOP: Average of estimated and theoretical standard errors with all values considered. MID: Average of estimated and theoretical standard errors with the largest outliers removed. BOTTOM: Average of estimated and theoretical standard errors with top two outliers removed.

though the discrepancy in the values is greatly reduced, there remains a relatively large difference between estimated and theoretical *se* values. Because of this, the lower part of Table 3 presents the averages with the two largest outliers removed from the calculations.
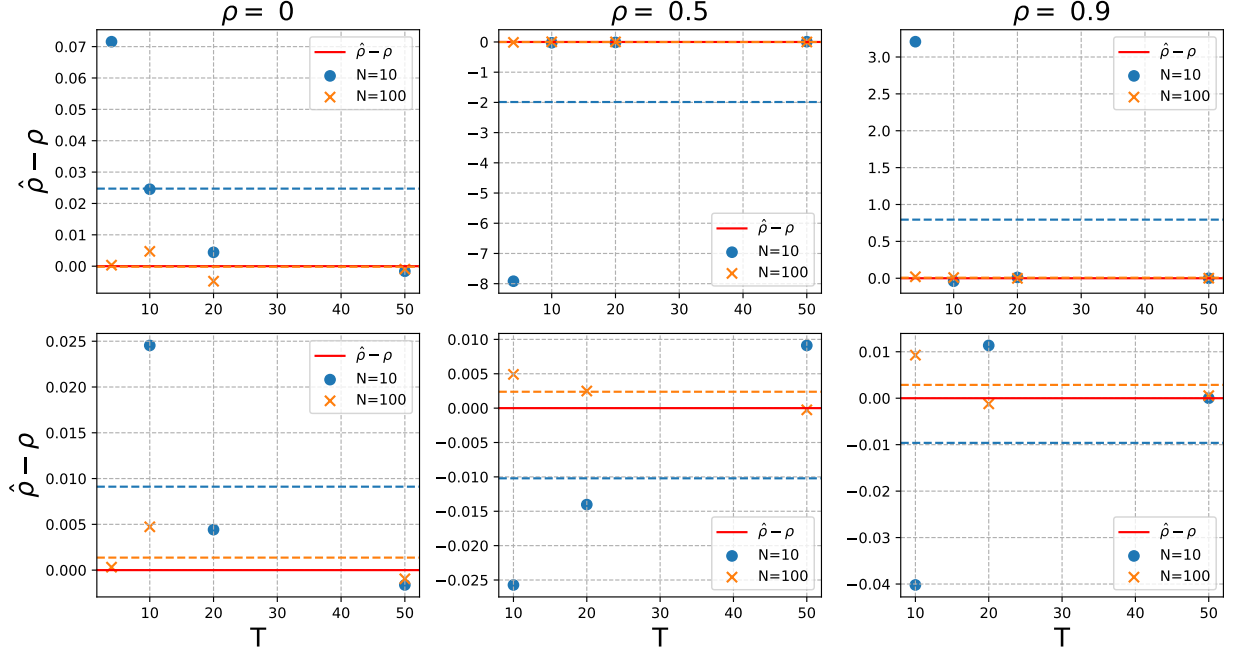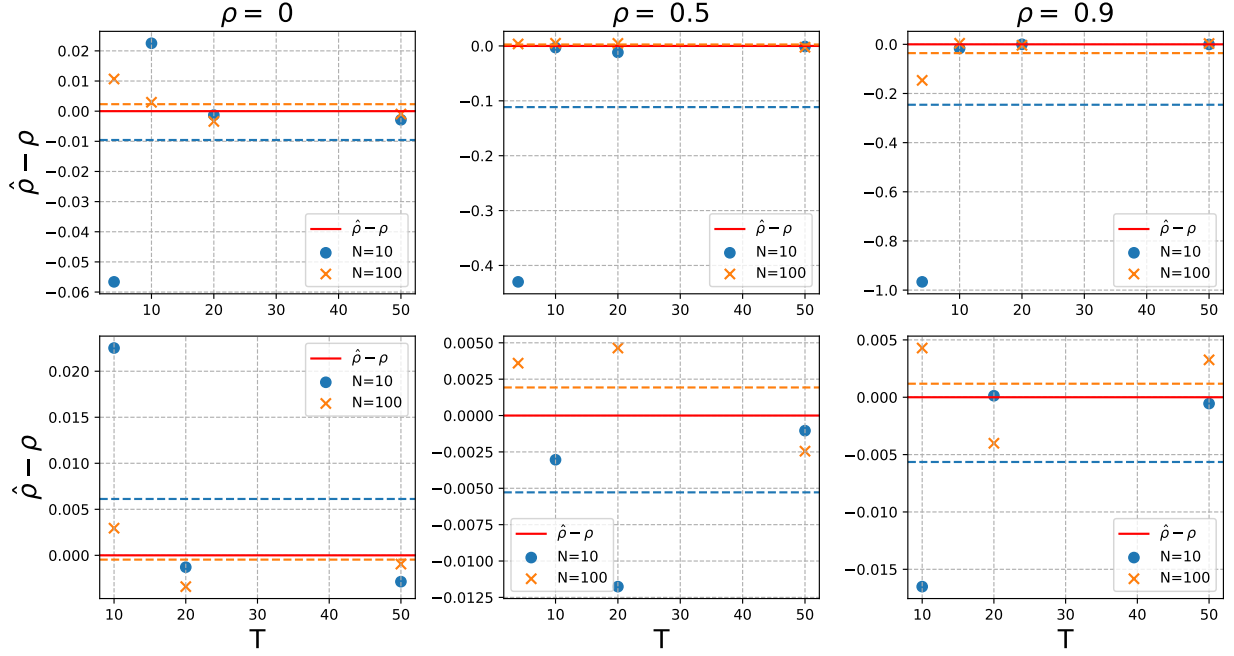
Figure 5: Difference between the estimator $\hat{\rho}_{GMM}$ and the true parameter ρ for the one-step GMM estimator. TOP: Values for all $\hat{\rho} - \rho$ estimated. BOTTOM: Values for $\hat{\rho} - \rho$ with the largest outliers removed

Figure 6: Difference between the estimator $\hat{\rho}_{GMM}$ and the true parameter ρ for the two-step GMM estimator. TOP: Values for all $\hat{\rho} - \rho$ estimated. BOTTOM: Values for $\hat{\rho} - \rho$ with the largest outliers removed
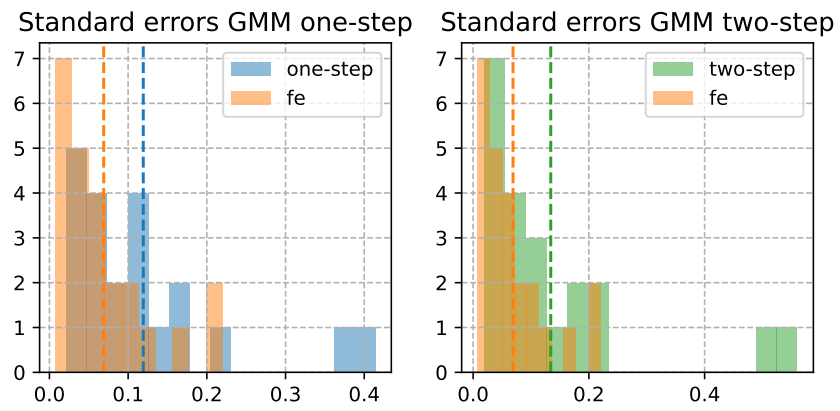
Figure 7: Standard errors of one-step GMM (left) and two-step GMM (right), compared to FE.