МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
СУМСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ

Кафедра прикладної математики та моделювання складних систем

Допущено до захисту
Завідувач кафедри ПМ та МСС
_____ Коплик І.В.

«___» _____ 2020 р.

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня «магістр»

спеціальність 113 «Прикладна математика»

освітньо-професійна програма «Прикладна математика»

тема роботи:

«МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ ПАНЕЛЬНИХ РЯДІВ СПОЖИВАННЯ ЕЛЕКТРОЕНЕРГІЇ»

Виконавець
студент факультету ЕлІТ
Смоленко Станіслав В’ячеславович
_____

Науковий керівник
ст. викл., к.е.н.
Маринич Тетяна Олександрівна
_____

Суми – 2020 р.

РЕФЕРАТ

Магістерська робота складається з 85 сторінок, містить 30 рисунків та 21 посилання на літературу.

Мета роботи – дослідження методів прогнозування часових та панельних рядів споживання електричної енергії на мікрорівні, а також вивчення профілю споживання електроенергії вищих навчальних закладів.

Об'єкт дослідження – дані споживання електричної енергії корпусами університету з автоматизованих систем комерційного обліку електроенергії.

Предмет дослідження – розробка панельних моделей та прогнозування короткострокової і довгострокової динаміки електроспоживання вищих навчальних закладів.

Актуальність налагодження системи моніторингу та прогнозування споживання енергоресурсів обумовлена чинниками енергобезпеки, енергоефективності, оптимізації фінансових ресурсів та зменшенням негативного впливу на довкілля.

В роботі досліджено економетричні методи аналізу та моделювання панельних рядів, запропоновано алгоритми практичної реалізації системи обліку, моніторингу та прогнозування споживання електроенергії. Досліджено різні фактори, які впливають на динаміку показників та їх відмінності за структурними підрозділами. Підготовлено рекомендації щодо використання прогнозної аналітики для прийняття ефективних управлінських рішень.

ЗМІСТ

ВСТУП

Україна посідає одне з найнижчих місць у світі за рівнем енергоефективності та енергоємності. Незважаючи на суттєве збільшення рівня ВВП на одиницю енергоспоживання [1] у останні десятиліття, показники споживання електроенергії на душу населення не змінилися з 1996 року [2]. Масове встановлення приладів обліку та контролю обсягів споживання електроенергії, використання енергоефективних пристроїв та технологій має стати відправною точкою у вирішенні проблеми. Прогнозування електроспоживання на мікрорівні, визначення факторів, які на нього впливають, а також напряму та сили ефекту зв’язків, дозволить підвищити результатівність початкових заходів.

Актуальність налагодження системи моніторингу та прогнозування споживання енергоресурсів обумовлена чинниками енергобезпеки, енергоефективності, оптимізації фінансових ресурсів та зменшенням негативного впливу на довкілля.

Метою роботи є дослідження методів прогнозування часових та панельних рядів споживання електричної енергії на мікрорівні, а також вивчення профілю споживання електроенергії вищих навчальних закладів.

Об’єктом дослідження є дані споживання електричної енергії корпусами університету з автоматизованих систем комерційного обліку електроенергії.

Предметом дослідження є розробка панельних моделей та прогнозування короткострокової і довгострокової динаміки електроспоживання вищих навчальних закладів.

1. ТЕОРЕТИЧНІ ТА МЕТОДОЛОГІЧНІ ПІДХОДИ МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ СПОЖИВАННЯ ЕЛЕКТРОЕНЕРГІЇ

1.1 Сучасні тенденції у прогнозуванні споживання електроенергії

Впровадження інтегрованих автоматизованих систем обліку та управління споживанням енергії є життєво важливим завданням як для світової економіки в цілому, так і для окремих країн, підприємств та організацій.

Розробка та валідація алгоритмів вимірювання, аналізу, моделювання та прогнозування споживання електроенергії покликана сприяти вирішенню таких нагальних проблем, як раціональне використання та диверсифікація ресурсів, забезпечення енергетичної безпеки та незалежності, енергоефективності, зменшення викидів парникових газів

Бібліометричний аналіз публікацій, проіндексованих базою даних Scopus [3], демонструє стійке зростання публікацій у галузі прогнозування споживання електроенергії (рис. 1.1).

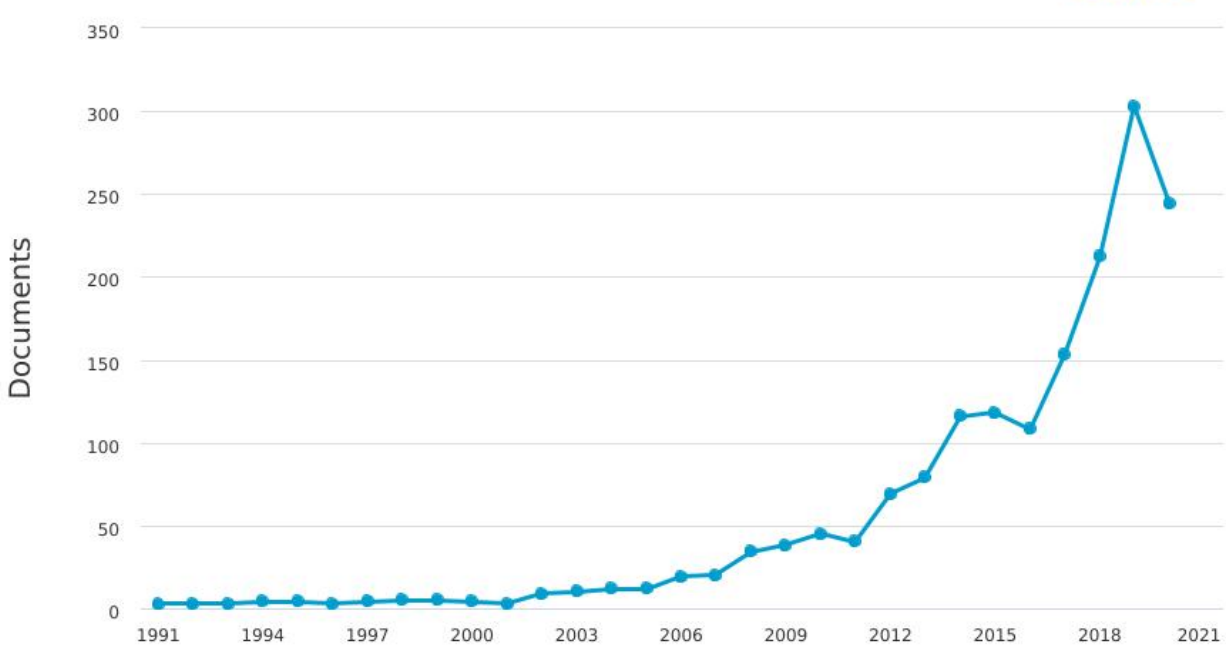


Рис. 1.1 – Динаміка публікацій з прогнозування електроспоживання в БД Scopus

При чому найбільше досліджень, представлених у [3], за період з 1990 по 2020 роки [3], опубліковано у виданнях з інженерії, комп'ютерних наук, енергетики, екології, математики та економіки (рис. 1.2).

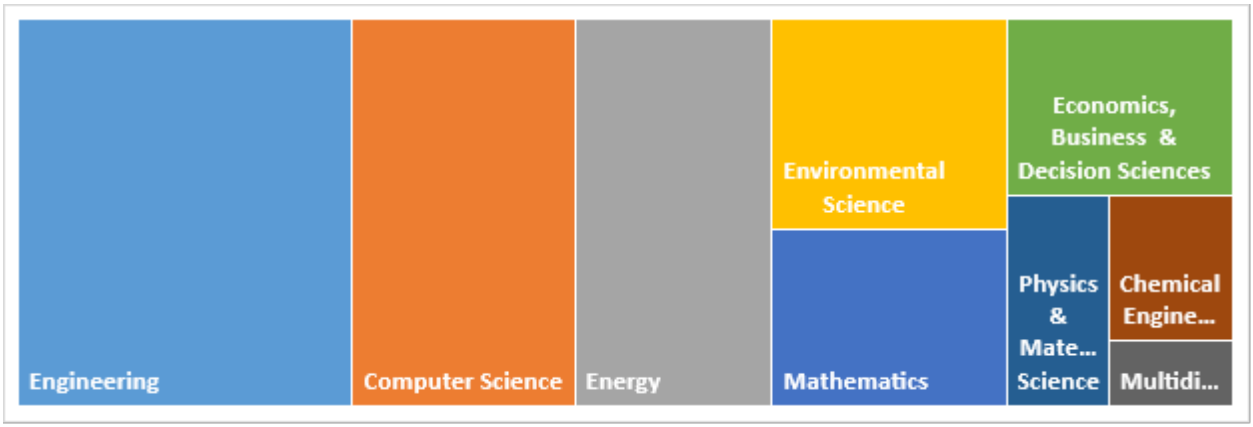


Рис. 1.2 – Структура публікацій БД Scopus з прогнозування електроспоживання за галузями знань

Зазначені дослідження проводилися переважно науковцями та організаціями з Китаю, США та Великобританії [3] (рис. 1.3).

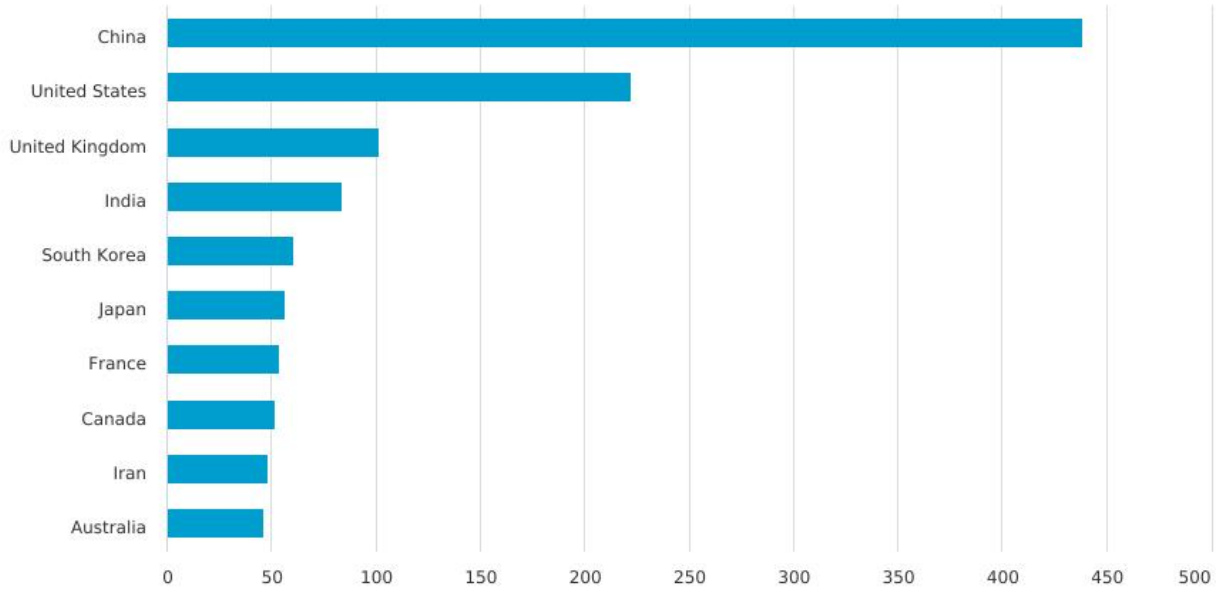


Рис. 1.3 – Структура публікацій БД Scopus з прогнозування електроспоживання у розрізі країн

Впровадження сучасних технологічних приладів вимірювання обсягів споживання електроенергії сприяло активізації розробки методів інженерного та статистичного аналізу для ефективного планування, прогнозування та контролю навантаження на електромережу. У розвинутих країнах поштовхом для масового впровадження спеціалізованого комп'ютерного програмного забезпечення стало державне регулювання енергоефективного проектування різних типів будівель, контролю за споживанням електроенергії та викидами CO2 [4].

Останнім часом подібні програмно-апаратні продукти з’являються і на українському ринку. Їх головною метою є підвищення енергоефективності та енергоощадності, а основне призначення полягає в організації системи моніторингу та управління електроспоживанням складних промислових систем, населених пунктів, групи будинків та ін. [5]. Одним з прикладів таких програм є автоматизована система комерційного обліку електроенергії (АСКОЕ) [6]. Аналогічна, проте менш вартісна, розробка науковців СумДУ [5] представляє систему з датчиків струму, мультиплексорів, мікроконтролерів опитування датчиків, мікроконтролеру зв’язку з сервером. Система в режимі реального часу перетворює показники струму з датчиків у електричну потужність за визначений період часу, далі передає аналоговий сигнал на контролер через мультиплексор, який виконує перемикання між 16-ма датчикам на канал АЦП мікроконтролера, і далі через WI-FI на сервер [5].

Підтримання енергоефективності в будівлях вимагає моніторингу в режимі реального часу показників споживання енергії та виявлення факторів, які впливають на них у реальному часі. При цьому важливо оцінювати як витрати електроенергії для виконання технологічного процесу (роботи об’єкта за призначенням), так і електроенергію, яка витрачається на компенсацію впливу зовнішніх факторів.

Більшість дослідників вважають погодні умови (температура повітря, тиск, швидкість та напрям вітру, хмарність, опади, тривалість сонячного дня)

основними чинниками зміни попиту на електроенергію. Також у моделі включають змінні постійного та змінного електричного навантаження; теплопостачання; календарні змінні; експлуатаційні характеристики будівель, наявність датчиків руху, класифікація комп'ютерного та іншого обладнання за типом енергоспоживання, параметри міської інфраструктури тощо [4].

Показники споживання електроенергії у навчальних закладах значною мірою залежать від розкладу занять, обладнання, яке використовується для лекційних та лабораторних занять, форм навчання – очної, заочної, дистанційної, змішаної чи гібридної. Виконання господарських договорів, проведення наукових досліджень та обслуговування допоміжних процесів (гуртки, секції, власні цехи з виробництва та обслуговування меблів і техніки), також впливає на обсяги споживання електроенергії і потребує окремого планування.

Якщо на початку 21 століття ще активно використовувалися інженерні моделі обліку та прогнозування попиту на електроенергію, то в останні роки все активніше використовуються статистичні моделі. Цьому сприяють збільшення потоків та масивів даних, розвиток програмно-апаратних комплексів для їх вимірювання та зберігання, формування ринку електроенергії та ринкового механізму визначення ціни на ресурси.

В літературі, для моделювання попиту на електроенергію в житловому, бюджетному та комунальному секторах використовують два статистичні підходи: "зверху вниз" і "знизу вгору" [6]. Перший підхід фокусується на визначенні ключових факторів та прогнозуванні споживання електроенергії житловими об'єктами різного рівня залежно від історичних даних сектору та змінних верхнього рівня, які включають макроекономічні показники (валовий внутрішній продукт, безробіття та інфляція), ціни на енергетичні ресурси, кліматичні фактори. Другий підхід заснований на моделюванні статистичних даних споживання електроенергії на регіональному та національному рівнях з

подальшою екстраполяцією на показники індивідуальних житлових комплексів [6].

В залежності від доступності частоти даних (щохвилинні, погодинні, тижневі, місячні, квартальні, річні) та мети дослідження автори виділяють моделі для короткострокових, середньострокових та довгострокових прогнозів. Для вирішення проблеми моделювання реальних статистичних даних, представлених різними частотами, застосовуються моделі змішаних частот, такі як (MIDAS) [5].

Крім того, поширеною є практика використання порогових значень показників (benchmarking), які є базою для порівняння фактичних і запланованих показників, аналізу відхилення від усереднених показників у галузі.

Ще одною проблемою при моделюванні споживання електричної та теплоенергії є обмеженість історичних даних. Літературний огляд показує, що для отримання адекватних статистично значущих та якісних прогнозних результатів за малої вибірки даних, дослідники часто застосовують підходи панельних моделей. Такі моделі дають оцінки та прогнози для груп подібних об’єктів, наприклад, для корпусів навчального закладу, житлових будинків регіону, регіонів країни або країн зі схожими параметрами розвитку. Різноманітність статистичних моделей зумовлена як різницею в структурі та типах даних, так і стрімким розвитком методів машинного навчання та програмних засобів для їх реалізації.

1.2 Економетричні моделі попиту на електроенергію

Оскільки історичні дані споживання електроенергії представлені у хронологічній послідовності за певний період, для їх прогнозування використовують переважно параметричні та непараметричні моделі часових рядів. Класичними прикладами параметричних моделей, які передбачають оцінювання параметрів за визначеного характеру розподілу даних, є лінійна та нелінійна регресії з часовими параметрами, авторегресійні моделі, моделі експоненційного згладжування, моделі Фур'є. Коли розподіл даних невідомий та / або коли розмір вибірки достатньо великий для навчання, використовують переважно непараметричні ансамблеві методи, такі як нейронні мережі, моделі градієнтного бустінгу, нечіткої логіки, моделі множинної сезонності TBATS, адитивні моделі нелінійного тренду та множинної сезонності Prophet тощо. Істотний недолік і обмеження непараметричних моделей полягає в складності їх оцінювання та інтерпретації, а також високих вимог до програмного та апаратного забезпечення [4].

В умовах обмеженості історичних даних, невиконання припущень класичних статистичних методів та необхідності визначення схожості або різниці між спорідненими об’єктами часто використовують панельні моделі.

На відміну від крос-секційних даних (*cross-section*), представлених спостереженнями за *n* об’єктами (суб’єктами), або часовими рядами (*time series*), представленими спостереженнями за *t* часовими періодами, панельні дані мають спостереження за n об’єктами у $T > 2$ часові періоди, що позначається як:

$$(X_{it}, X_{it}), i = 1, \dots, n \ \ and \ \ t = 1, \dots, T.$$

Більшість методів для панельних даних передбачають попереднє збалансування даних, коли кожен об’єкт (*cross-section*) має однакову кількість спостережень у часі (*time series*). Проблема пропущених даних у деяких

періодах вирішується інтерполяцією даних в межах окремих об’єктів або заповненням попередніми чи наступними значеннями.

Регресійні моделі з використанням панельних даних вважаються дієвим економетричним методом, що забезпечує більшу інформативність, більшу ефективність оцінок в умовах значної змінності та неоднорідності даних, а також пом'якшення проблеми пропущених та мультиколінеарних змінних [9].

Загальний вигляд лінійної панельної регресії представлено у (1.1):

$$Y_{it} = \alpha + \beta X_{it} + u_{it} \tag{1.1}$$

Тут змінні *Y* та *X* мають обидва *i* та *t* нижні індекси для *i = 1, 2, ..., N* об’єктів (*cross-sections*) і *t = 1, 2, ..., T* часових періодів. Коефіцієнти *α* та *β* не мають нижніх індексів, припускаючи, що вони будуть незмінними для всіх об’єктів протягом всього досліджуваного періоду.

Оскільки передумови методу найменших квадратів (МНК), що використовується для оцінки параметрів лінійної регресійної моделі, не виконуються, його не можна застосовувати для оцінки панельних моделей.

Загальний вигляд лінійної регресійної моделі представлено у (1.2):

$$Y_i = \alpha + \beta X_i + \epsilon_i \tag{1.2}$$

МНК-оцінки коефіцієнтів регресії знаходяться як:

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{1.3}$$

МНК-оцінки параметрів моделі (1.3) є незміщеними $E(\hat{\beta}) = \beta$, відповідними $(lim_{n\to\infty}\hat{\beta}) = \beta$ та ефективними $Var(\hat{\beta}) \leq Var(\tilde{\beta})$, якщо виконуються такі передумови:

- Гомогенність залишків: $Var(\epsilon_i) = \sigma^2 \; for \; i = 1,2, \dots, n.$
- Відсутність автокореляції залишків: $Cov(\epsilon_i \epsilon_j) = 0 \; for \; i \neq j.$
- Екзогенність: $E(\epsilon_i | x_i) = 0 \; for \; i \neq j.$

Невиконання передумов МНК призводить до:

1. Гетероскедастичність залишків => неефективність $\hat{\beta}$.

2. Серійна кореляція (автокореляція) залишків => неефективність $\hat{\beta}$.
3. Ендогенність => невідповідність $\hat{\beta}$.

До методів, які можуть бути застосовані, у разі невиконання зазначених вище передумов відносять:

1) метод узагальнених найменших квадратів для моделей з гетероскедастичністю та серійною кореляцією (Generalized Least Squares, GLS);
2) методи інструментальних змінних: двокроковий метод найменших квадратів; узагальнений метод моментів (Generalized Method of Moments, GMM) для моделей з ендогенністю.

В цілому, лінійні панельні моделі можуть бути оцінені, використовуючи три різні методи [9]:

1) із спільною константою α у рівнянні (також має назву складений або пуловий метод найменших квадратів, *the pooled OLS method*):

$$Y_{it} = \alpha + X_{it}\beta + \epsilon_{it}, \ E(\epsilon|X) = 0 \quad (1.4)$$

2) метод, що дозволяє фіксовані ефекти (*the fixed effects method*):

$$Y_{it} = \alpha_i + X_{it}\beta + \epsilon_{it}, \quad (1.5)$$

де α_i представляють індивідуальні ефекти $n-1$ об'єктів, які корелюють з X_{it};

3) метод, оснований на випадкових ефектах (*the random effects method*).

$$Y_{it} = \alpha_i + X_{it}\beta + \epsilon_{it}, \quad (1.6)$$

$$Y_{it} = \alpha + X_{it}\beta + (\nu_i + \epsilon_{it}), \quad (1.7)$$

де $\alpha_i = \alpha + \nu_i$; ν_i – це стандартизована випадкова величина з нульовим математичним сподіванням та сталою дисперсією (μ=0; σ_ν^2).

Модель з фіксованими ефектами може бути оцінена двома способами:

1) МНК з фіктивними змінними для $n-1$ об'єктів (Least Squares Dummy Variable, LSDV, або Within Estimator):

$$Y_{it} - \bar{Y}_i = \alpha_i + X'_{it}\beta + \epsilon_{it} - (a_i + \bar{X}'_i\beta + \bar{\epsilon}_i) =$$

$$=(X_{it} - \bar{X}_i)'\beta + (\epsilon_{it} - \bar{\epsilon}_i) \qquad (1.8)$$

2) МНК оцінка перших різниць:

$$Y_{it} - Y_{i,t-1} =(X_{it} - X_{i,t-1})'\beta + (\epsilon_{it} - \epsilon_{i,t-1}) \qquad (1.9)$$

Модель з фіксованими ефектами з фіктивним змінними може бути записана у матричній формі як:

$$Y = D\alpha + X\beta' + u \qquad (1.10)$$

Тут маємо:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_N \end{pmatrix}_{NT\times 1}, \; D = \begin{pmatrix} i_T & 0 & \dots & 0 \\ 0 & i_T & & 0 \\ \dots & \dots & & \dots \\ 0 & 0 & & i_T \end{pmatrix}_{NT\times N}, \; X = \begin{pmatrix} x_{11} & x_{21} & .. & x_{1k} \\ x_{21} & x_{22} & & x_{2k} \\ \dots & \dots & & \\ x_{N1} & x_{N2} & & x_{Nk} \end{pmatrix}_{NT\times k} \quad \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_N \end{pmatrix} \; \beta' = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{pmatrix} \qquad (1.11)$$

Відмінність між двома способами оцінювання панельних даних – з фіксованими (1.5) або випадковими ефектами (1.6) – полягає в тому, що у (1.5) кожен об'єкт відрізняється від інших на величину константи α_i, тоді як у (1.6) об'єкти відрізняються на величину збурення v_i. Недоліком моделі з випадковими ефектами є те, що вона містить припущення про нормальний розподіл випадкової компоненти v_i, що на практиці не завжди можливо. Проте, оскільки вона має менше невідомих параметрів, ніж модель з фіксованими ефектами, це дозволяє включати додаткові пояснювальні змінні в модель (фіктивні змінні, притаманні усім об'єктам).

Для визначення необхідності включення фіксованих ефектів (різних констант для різних крос-секцій) до регресійної моделі порівняно зі сталою константою у пуловому методі використовують стандартний *F* тест [9], де

нульова гіпотеза засвідчує, що всі константи однакові і, відповідно пуловий метод із спільною константою має застосовуватися:

$$H_0 : \alpha_1 = \alpha_2 = ... = \alpha_N \tag{1.12}$$

Для визначення моделі, яка краще описує експериментальні статистичні дані, застосовують тест Хаусмана (*Hausman, 1978*) [9], нульова гіпотеза H_0 якого полягає у тому, що випадкові ефекти є послідовними та ефективними проти альтернативної гіпотези H_1, що фіксовані ефекти дають незміщені ефективні оцінки параметрів панельної регресії.

Для визначення пріоритетності між панельною моделлю з фіксованими ефектами та пуловою моделлю зі спільною константою застосовують тест Вальда (*Wald Test*) [9], який перевіряє значущість (відмінність від нуля) коефіцієнтів регресії: H_0: $c(1)=c(2)=0$.

Перевірка доцільності моделі з випадковими ефектами здійснюється на підставі тесту *Breusch-Pagan* – H_0: відсутні випадкові ефекти проти H_1: випадкові ефекти присутні [9].

Алгоритм відбору регресійної моделі для панельних даних представлена на рис. 1.4.

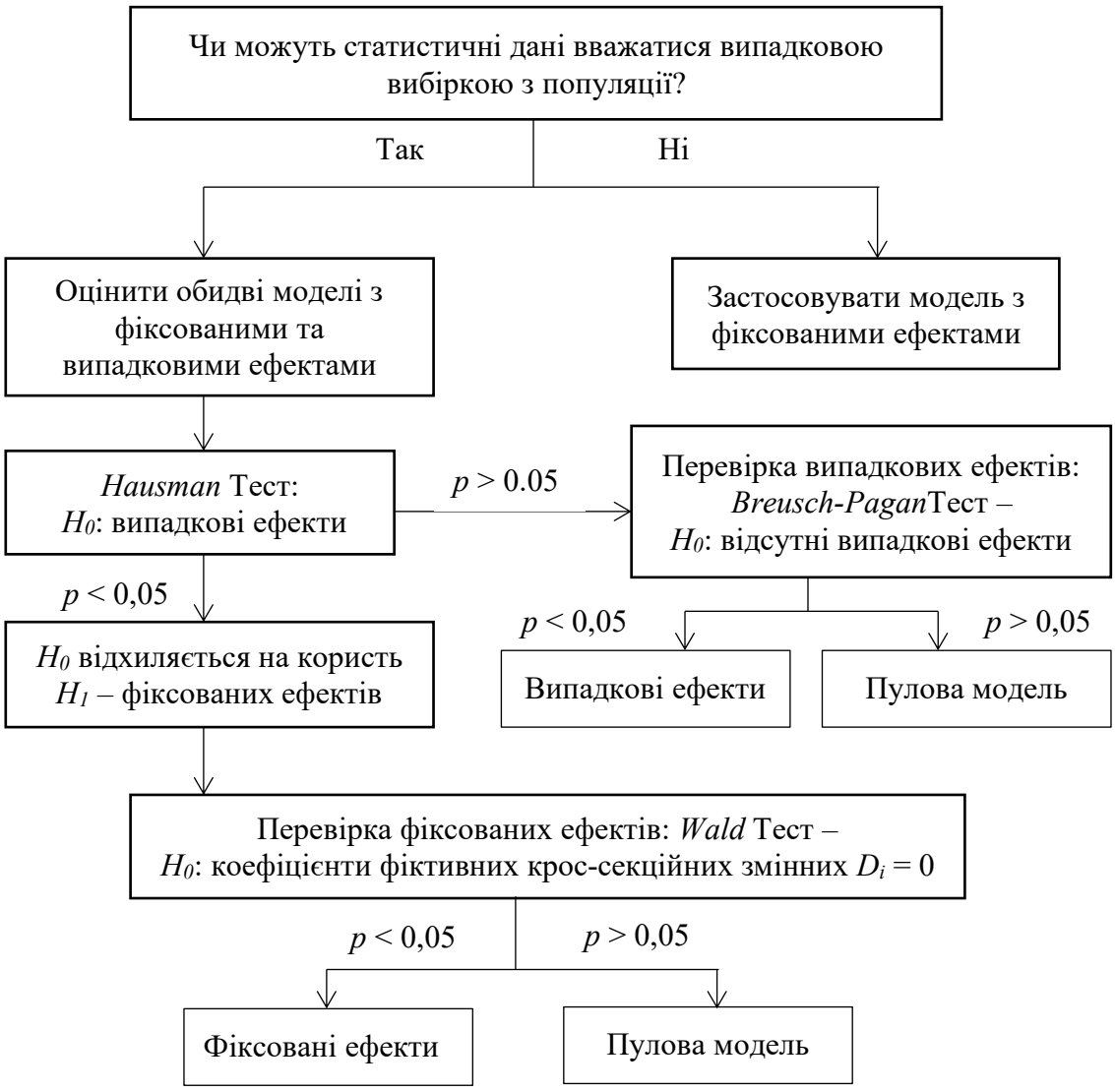


Рис. 1.4 – Алгоритм відбору регресійної моделі для панельних даних

Умовою отримання адекватних оцінок параметрів регресійних моделей є стаціонарність пулових рядів та індивідуальних часових рядів об’єктів. Тому алгоритм панельних моделей передбачає тестування змінних на наявність одиничних коренів (*Unit Root Test*), спричинених систематичними компонентами (трендом, сезонністю, циклічністю), які порушують умови стаціонарності: μ=*const*, $\sigma^2 = const$, $Cov(y_i y_j) = 0$ [9].

Прикладами таких тестів, адаптованих для панельних даних, є розширений тест Левіна, Ліна та Чу (Levin, Lin, and Chu, LLC) та тест Іма, Песарана та Шіна (Im, Pesaran and Shin, IPS), які передбачають наявність

загальних та індивідуальних авторегресійних параметрів за крос-секціями. Обидва тести побудовані на базі розширеного тесту Дікі-Фулєра (Dickey-Fuller, ADF) на наявність одиничних коренів [10]:

$$\Delta y_{it} = \alpha y_{it-1} + \sum_{j=1}^{p_i} \beta_{ij} \Delta y_{it-j} + X'_{it}\delta + \epsilon_{it} \quad (1.13)$$

Тут y_{it} є панельним рядом, $\Delta y_{it} = y_{it} - y_{it-j}$, $\alpha = \rho - 1$ є загальним авторегресійним коефіцієнтом, p_i є порядком лагу для різниць ряду Δy_{it}, який може змінюватися в залежності від крос-секцій.

Тест LLC перевіряє наявність загальних одиничних коренів: нульова гіпотеза $H_0: \alpha = 0,$ альтернативна $H_1: \alpha < 0$. Тест IPS оцінює індивідуальні (за крос-секціями) процеси одиничних коренів та перевіряє стаціонарність ряду в умовах крос-секційної залежності (кореляції між об'єктами):

$$H_0: \alpha_i = 0 \; for \; all \; i \quad (1.14)$$

$$H_1: \alpha_i = 0 \; for \; i = 1,2, \dots, N_1, \; H_1: \alpha_i < 0 \; for \; i = N+1, N+2, \dots, N$$

Для приведення рядів до стаціонарності застосовується процедура диференціювання (від кожного наступного значення віднімається попереднє значення крос-секційних рядів) – $\Delta y_{it} = y_{it} - y_{it-j}$.

Оскільки традиційні оцінки параметрів панельних моделей з фіксованими та випадковими ефектами можуть бути невідповідними або неефективними, коли панельна вибірка неоднорідна, та існують загальні шоки та неспостережувані компоненти, необхідно додатково перевірити крос-секційну залежність (CD) у помилках [11]. Так, наприклад, Pesaran CD тест, базується на наступній статистиці CD:

$$CD = \sqrt{\frac{2T}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{(T-k)\hat{\rho}_{ij}^2 - E[(T-k)\hat{\rho}_{ij}^2]}{var\left[(T-k)\hat{\rho}_{ij}^2\right]} \quad (1.15)$$

де $\hat{\rho}_{ij}$ є кореляцією між кожною парою залишків МНК.

Щоб врахувати серійну кореляцію помилок всередині групи, неоднорідність панельної вибірки, а також короткострокові та довгострокові

ефекти незалежних змінних на залежну змінну, поширеною практикою є використання панельних моделей авторегресії розподіленого лагу (Panel Auto-Regressive and Distributed Lag Model, ARDL). Даний підхід дає можливість вивчити динамічні причинно-наслідкові зв'язки [12]. Специфікацію узагальненої моделі ARDL можна записати наступним чином:

$$y_{it} = \sum_{j=1}^{p} \delta_i y_{it-j} + \sum_{j=0}^{q} \beta'_{ij} \Delta x_{it-j} + \varphi_i + \gamma_t + \epsilon_{it} \qquad (1.16)$$

Тут y_{it} – це стаціонарна залежна змінна. x_{it} – це $k \times 1$ вектор пояснювальних стаціонарних змінних. δ_iта β'_{ij} – коефіцієнти лагових змінних y_{it} та x_{it}.

φ_i – групові фіксовані ефекти; γ_t – періодичні специфічні ефекти (фіксовані або випадкові). p, q – оптимальні порядки лагів y_{it} та x_{it} згідно з інформаційними критеріями Акакайке (AIC) та Шварца (SC). ϵ_{it} – це залишковий компонент. $\sim N(0, \sigma^2)$.

Для оцінювання короткострокової та довгострокової динаміки змінних використовують репараметризовану модель ARDL:

$$\Delta y_{it} = \theta_i [y_{it-1} - \lambda'_i x_{it}] +$$

$$+ \sum_{j=1}^{p-1} \xi_{ij} \Delta y_{it-j} + \sum_{j=0}^{q-1} \beta'_{ij} \Delta x_{it-j} + \varphi_i + \gamma_t + \epsilon_{it} \qquad (1.17)$$

Тут Δy_{it}, Δx_{it} – перші різниці змінних, що приводять дані до стаціонарності. $\theta_i = -(1 - \delta_i)$ індивідуальним для кожної групи коефіцієнтом пристосування до довгострокової рівноваги. λ'_i = вектор довгострокового зв'язку. $ECT = [y_{it-1} - \lambda'_i x_{it}]$ рівняння довгострокової корекції. ξ_{ij}, β'_{ij} - коефіцієнти короткострокової динаміки.

Виділяють такі методи оцінки ARDL моделей:

- Середньогруповий Mean Group (MG);
- Пуловий середньогруповий метод (Pooled Mean Group, PMG);
- Динамічний МНК з фіксованими ефектами (Dynamic Fixed Effects Ordinary Least Squares, DFE OLS);

- Узвгальнений метод моментів (Generalized Method of Moments Estimator, GMM).

Зазначені вище оцінки вважаються робастними оцінками довгострокових та короткострокових параметрів у присутності крос-секційної залежності (CD), ендогенності, автокореляції [12], [13].

Алгоритм методу GMM передбачає:

1) Обрахунок перших різниць:

$$(Y_{it} - Y_{i,t-1}) = \gamma(Y_{i,t-1} - Y_{i,t-2}) + (X_{it} - X_{i,t-1})\beta + (\epsilon_{it} - \epsilon_{i,t-1})$$

2) Використання $Y_{i,t-2}$ як інструментальної змінної (Z_i) та оцінювання параметрів за двокроковим МНК.
3) виконання GMM:

$$\hat{\beta}_{AB} = \left[\left(\sum_{i=1}^{N} \tilde{X}_i' Z_i\right) W_N \left(\sum_{i=1}^{N} Z_i' \tilde{X}_i\right)\right]^{-1} \left(\sum_{i=1}^{N} \tilde{X}_i' Z_i\right) W_N \left(\sum_{i=1}^{N} Z_i' \tilde{Y}_i\right)$$

Для визначення короткострокових причинно-наслідкових зв’язків панельних даних додатково використовують тест причинності або каузальності Думітреску-Хурліна (Dumitrescu-Hurlin test), який є видозміненим варіантом тесту Грейнджера (Granger test) для часових рядів [14]. Регресійна модель тесту виглядає наступним чином:

$$y_{it} = \alpha_i + \sum_{j=1}^{p} \beta_{ij} y_{it-j} + \sum_{j=0}^{q} \gamma_{ij} \Delta x_{it-j} + \epsilon_{it} \qquad (1.18)$$

Нульова гіпотеза тесту (1.18), що відповідає відсутності причинно-наслідкового зв’язку для всіх крос-секцій (об’єктів) у панелі. визначається як:

$$H_0: \gamma_{i1} = \cdots = \gamma_{ip} = 0 \quad \forall\, i = 1, \ldots, N \qquad (1.19)$$

Тест (1.18) припускає, що для деяких крос-секцій може бути причинно-наслідковий зв'язок, але не обов'язково для всіх. Таким чином, альтернативна гіпотеза має вигляд:

$$H_1: \gamma_{i1} = \cdots = \gamma_{ip} = 0 \quad \forall\, i = 1, \ldots, N_1$$

$$\gamma_{i1} \neq \cdots \neq \gamma_{ip} \neq 0 \quad \forall\, i = N_1 + 1, \ldots, N \tag{1.20}$$

Алгоритм тесту Думітреску-Хурліна передбачає оцінювання N індивідуальних регресій (1.18), виконання F тесту p лінійних гіпотез (1.19) для отримання W_i і обрахунку узагальненої для T періодів статистики $\bar{W}$ як середнє з N індивідуальних статистик Вальда (Wald statistics):

$$\bar{W} = \frac{1}{N}\sum_{i=1}^{N} W_i \tag{1.21}$$

2. ПІДГОТОВКА ТА ОПИС ДАНИХ

За допомогою програмного комплексу обліку споживання електроенергії АСКОЕ нами було завантажено часові ряди даних споживання електроенергії по дням за основними корпусами університету у період з 1 січня 2015 р. по 30 листопада 2020 р.. Показники електроспоживання представлені такими характеристиками:

- A+: споживання активної електроенергії за розрахунковий період, кВт * год.;
- R-: споживання реактивної електроенергії (перетікання реактивної електроенергії з мережі енергопостачальної організації в мережу споживача) за розрахунковий період, кВар.год..

Показники електроспоживання збираються за корпусами та гуртожитками Центрального кампусу університету. Деякі корпуса мають дві точки входу через збільшене навантаження за електропостачанням. Для подальшого аналізу ми будемо аналізувати узагальнені показники, обраховані як сума за усіма точками входу корпусу. Початковий період обліку даних за лічильниками також різниться – деякі корпуси мають показники з 1 січня 2015 року, деякі з 2017 або 2018 років. Тому ми розбили дані на дві вибірки – перша вибірка містить повні дані за 2015 – 10’2020 рр.; друга вибірка містить повні дані з 11’2017 по 10’2020 рр..

На рис. 2.1 – 2.3 представлена динаміка споживання активної електроенергії по центральному та головному корпусам університету за 2015-2020 рр. Як видно з графіків, дані споживання активної електроенергії містять аномальні спостереження на початку обрахунку показників за лічильниками (рис. 2.2 – 2.3) та усередині часових періодів (рис. 2.1, 2.3). Це пояснюється збоями в системі або проблемами у роботі лічильників. Аномально низькі дані лічильників, які не відповідають загальній динаміці споживання, було замінено інтерполяцією відповідних даних з попередніх періодів.

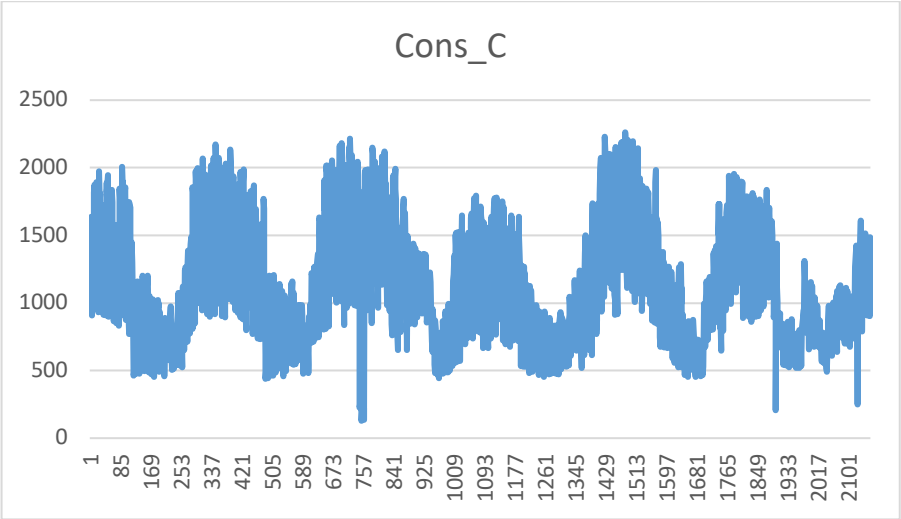


Рис. 2.1 – Показники електроспоживання по корпусу Ц за 01/01/2015-30/11/2020

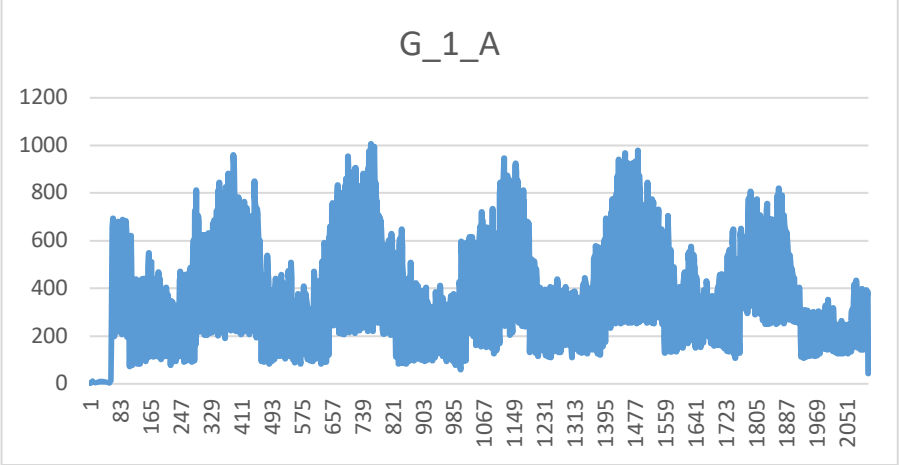


Рис. 2.2 – Показники електроспоживання по корпусу Г, точка вводу 1, за 01/01/2015-30/11/2020

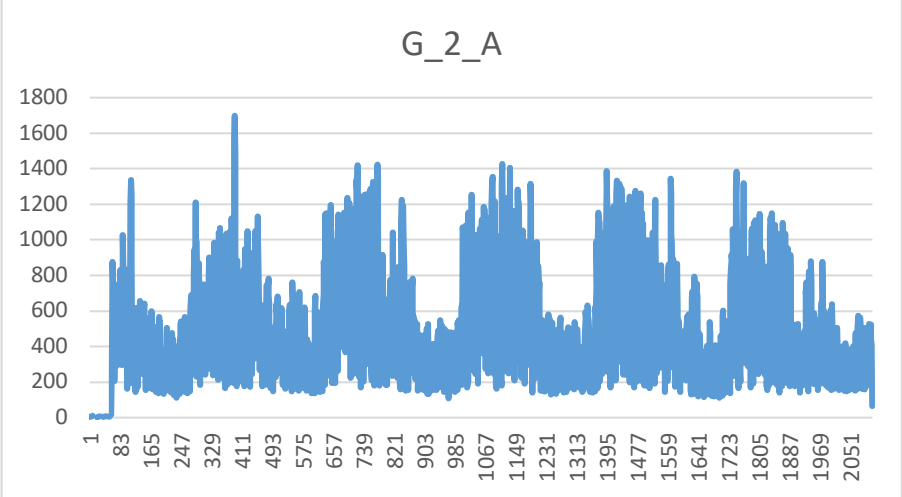


Рис. 2.3 – Показники електроспоживання по корпусу Г, точка вводу 2, за 01/01/2015-30/11/2020

Річна динаміка споживання активної та реактивної електроенергії корпусом НВК у 2015 р. представлена на рис. 2.4 – 2.5. Показники споживання реактивної електроенергії (рис. 2.5) надалі вивчатися не будуть, оскільки вони відображують технологічні процеси, які супроводжують активне сподивання.

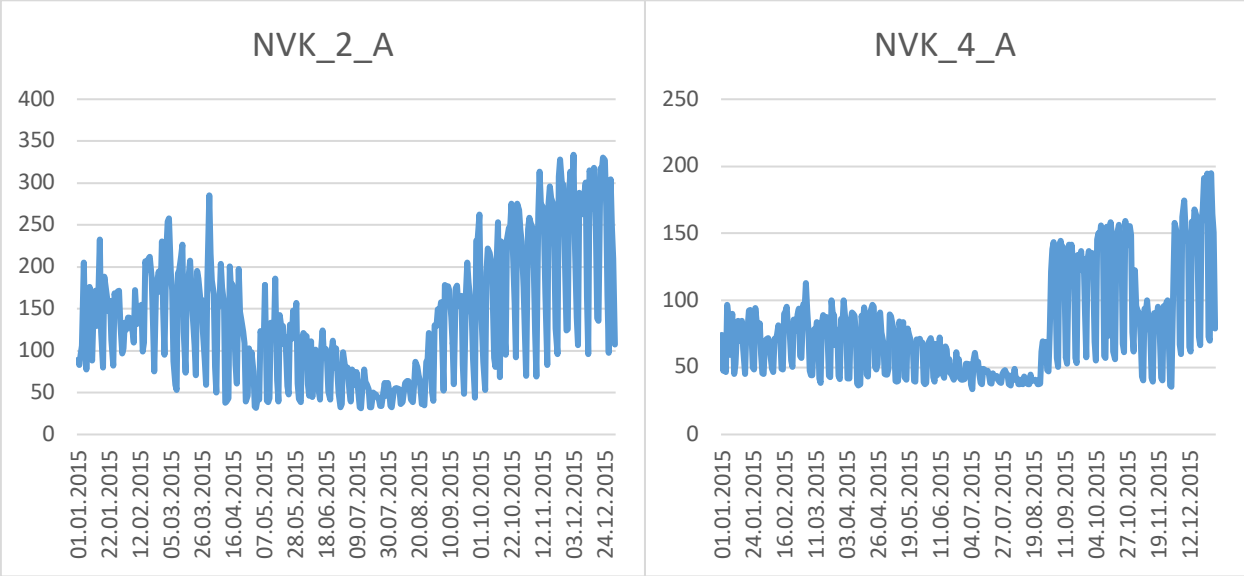


Рис. 2.4 – Показники споживання активної енергії (А+) у кВт*год.
за корпусом НВК (точки входу 2 та 4) у 2015 р.

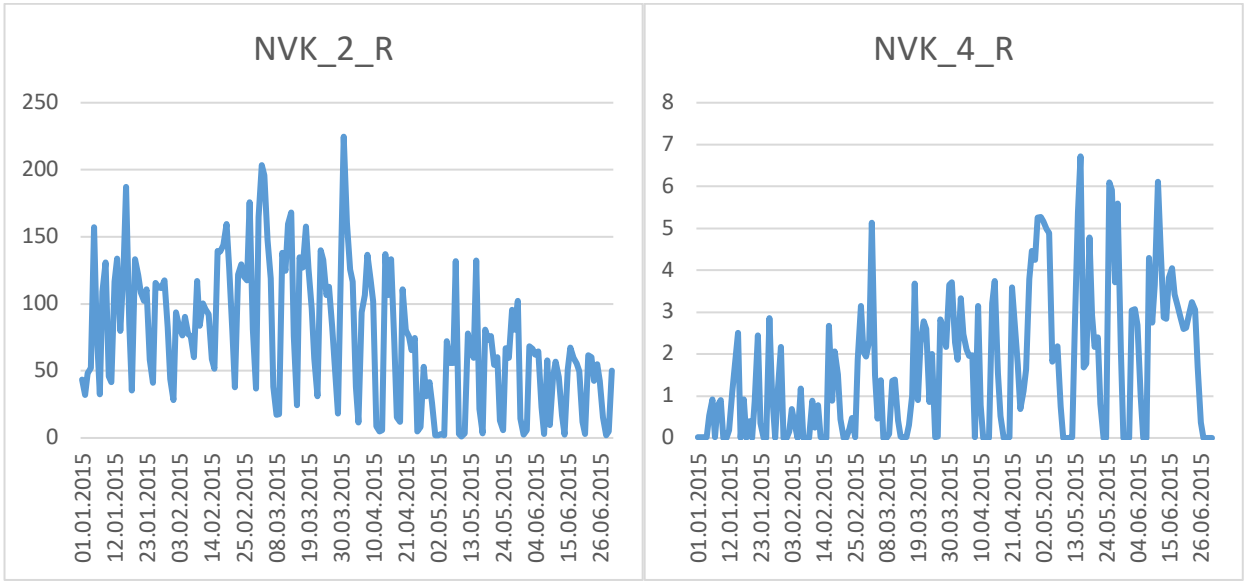


Рис. 2.5 – Показники споживання реактивної енергії (R-) у кВар.год.
за корпусом НВК у 2015 р.

Додатково у роботі досліджувалися погодні дані по м. Суми за період 2015 – 2020 роки [15]. Оскільки архівні погодні дані предсавлені щоденно через кожні три години, ми здійснили усереднення цих показників по дням.

Для формування панельної вибірки дані споживання електроенергії за всіма корпусами були додані до однієї змінної (Cons), виділена окрема змінна – назва корпусу (Object); дані погоди та календарні ефекти скопійовано відповідним чином для всіх корпусів. Зведена таблиця вхідних емпіричних даних представлена у таблиці 2.1.

Таблиця 2.1 Змінні емпіричної панельної вибірки

Змінна	**Значення**
year	Рік
month	Місяць
day	День
Date	Повна дата
Weekend	Фіктивна змінна, що приймає значення 1 для вихідних (субота, неділя) та 0 у інших випадках
Holiday	Фіктивна змінна = 1 для святкових днів та 0 у інших випадках
Object	Назва корпусу (G, C, NVK, M, Lib, L, T)
Cons	Щоденне споживання електроенергії, кВт*год
T	Середня щоденна температура повітря на висоті 2 метри над поверхнею землі, градуси Цельсія
p0	Середній за період атмосферний тиск на рівні станції, мм рт. ст.
p	Середній за період атмосферний тиск, наведений до середнього рівня моря, мм рт. ст.
Pa	Середня за період барична тенденція: зміна атмосферного тиску за останні три години, мм рт. ст.
U	Середня за період відносна вологість на висоті 2 метри над поверхнею землі, %
DD	Напрямок вітру на висоті 1-12 метрів над поверхнею землі
Ff	Середня швидкість вітру на висоті 10-12 м над поверхнею землі, м / с
RRR_a	Середня кількість опадів, мм
light	Тривалість світлого часу доби за період, частка від одиниці

Слід зазначити, що для формування повноцінної панельної вибірки, яка б описувала відповідні відмінності між корпусами, нам бракувало специфічних даних, властивих тому чи іншому корпусу, таких як:

- кількість людей, що відвідало підрозділ за добу;
- кількість та види електроприладів і пристроїв, що умикалися за добу;
- кількість проведених навчальних занять та їх видів.

Алгоритм моделювання та прогнозування споживання електроенергії передбачає такі етапи:

1. Збір даних для моделі (Рис. 2.6):

a. Дані електроспоживання з лічильників та датчиків.

b. Дані щодо запланованого споживання з наперед визначеними навантаженням.

c. Дані зміни погоди та календарні ефекти.

d. Показники енергоефективності за об’єктом.

2. Виділення фіксованої та випадкової компонент часових рядів електроспоживання.

3. Створення пулової вибірки даних з часових рядів корпусів.

4. Обробка та трансформація даних, робота з аномаліями.

5. Перевірка стаціонарності та крос-секційної залежності рядів.

6. Моделювання за класичними лінійними панельними моделями – пулові моделі, моделі з фіксованими та випадковими ефектами.

7. Моделювання за динамічними панельними моделями розподіленого лагу.

8. Перевірка статистичних та прогнозних властивостей моделей.

9. Прогнозування та анліз причинно-наслідкових звязків. Декомпозиція прогнозу відносно наперед визначеного фіксованого споживання та залишкового споживання.

10. Формулювання рішень для підвищення енергоефективності та енергозбереження.

Рис. 2.6 – Збір даних для моделі прогнозування електроспоживання

Вибірка даних за 2015 – 10’2020 рр. (*panel_2015*) містить дані за трьома корпусами – Ц (C), Г (G), НВК (NVK) і включає 6423 спостережень та 17 змінних.

Вибірка даних за 11’2017 – 10’2020 рр. (*panel*) містить дані за 7 корпусами – Центральний (C), Головний (G), НВК (NVK), Лабораторний (L), корпус М (M), корпус Т (T), корпус БІЦ (Lib) і включає 7742 спостережень та 17 змінних.

Додатково оцінювалися такі вибірки за 11’2017 – 10’2020 рр.:

- *panel_cg* – вибірка за корпусами «Ц» та «Г»;
- *panel_c* – вибірка по Центральному корпусу;
- *panel_g* – вибірка по Головному корпусу;
- *panel_s1*– вибірка по корпусам НВК, Л, М, Т, БІЦ.

Також для виявлення довгострокової причинності та побудови ARDL моделей оцінювалися вибірки за узагальненими місячними даними.

3. МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ ПАНЕЛЬНИХ ДАНИХ СПОЖИВАННЯ ЕЛЕКТРОЕНЕРГІЇ

3.1 Статистичний аналіз рядів споживання електроенергії

Початковий етап дослідження включає статистичний аналіз даних, візуалізацію розподілу змінних та взаємозв'язків між змінними.

Таблиця 3.1 представляє основні показники описової статистики змінної електроспоживання вибірки 11'2017 – 10'2020 рр., за всіма корпусами та за однорідними групами корпусів. Гістограми розподілу змінних *Cons* у розрізі різних корпусів та вихідних, святкових днів зображені на рис. 3.1 – 3.2.

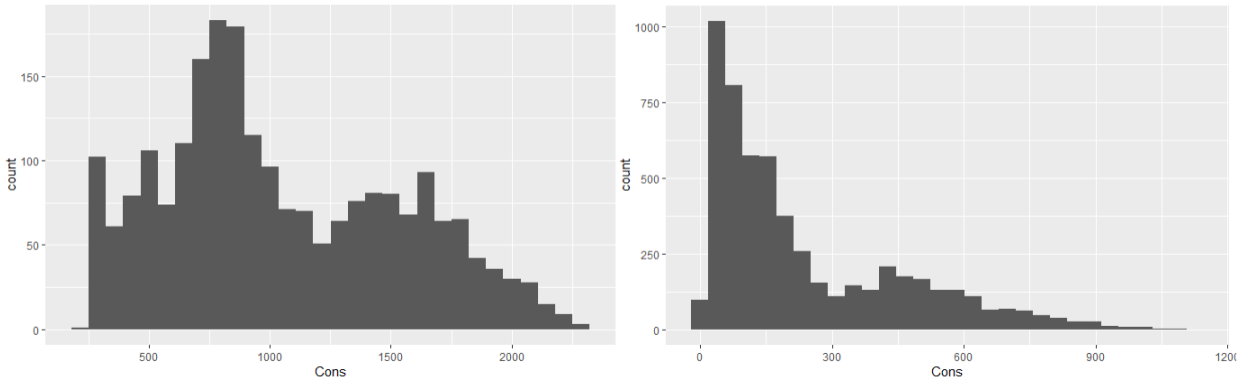
Як показує табл. 3.1, медіана ряду *Cons* за усім корпусами значно відрізняється від середнього значення, що означає, що розподіл ряду не є симетричним (нормальним), а зміщений вправо, у бік великих значень. З рис. 3.1-3.2 видно, що розподіли рядів споживання електроенергії є бімодальними, оскільки дані по робочім дням скупчені навколо одних показників центральної тенденції, а по вихідним – навколо інших.

Таблиця 3.1 Описова статистика ряду електроспоживання (*Cons*)

Статистика	***Cons*** **усі корпуси**	***Cons*** **корп. *C, G***	***Cons*** **корп. *L, M, T, NVK, Lib***
Мінім. значення	9.04	206.7	9.04
1 квартиль	97.77	687.5	69.22
Медіана (2 квартиль)	285.28	918.3	152.37
Середнє	469.55	1046.1	238.94
3 квартиль	689.63	1445.7	379.77
Максим. значення	2279.88	2279.88	1135.50
Ст. відхилення	485.30	488.56	219.38
N (кількість спостережень)	7742	2212	5530

(а) (б)

Рис. 3.1 – Гістограма розподілу споживання електроенергії (*Cons*) за 7 корпусами у 11’2017 – 10’2020 рр. у розрізі вихідних (а) та святкових (б) днів



(а) (б)

Рис. 3.2 – Гістограма розподілу споживання електроенергії (*Cons*) за корпусами *C, G* (а) та корпусами *L, M, T, NVK, Lib* (б) у 11’2017 – 10’2020 рр.

Відмінності у споживанні електроенергії у розрізі усіх корпусів Центрального кампусу університету продемонстровані на діаграмі *boxplot* (рис. 3.3.) [16]. Даний вид діаграми дає можливість побачити наочно різницю у мінімальному, максимальному значеннях, першому, другому(медіані) та третьому квантилях даних. Так, корпуси *C, G* мають найбільші показники споживання електроенергії, а корпуси *M, T* – найменші.

Крім того, можемо побачити відмінності у споживанні електроенергії у розрізі місяців (рис. 3.4.). Закономірно, що найбільше споживання зафіксовано у січні, лютому, жовтні, листопаді та грудні місяцях.

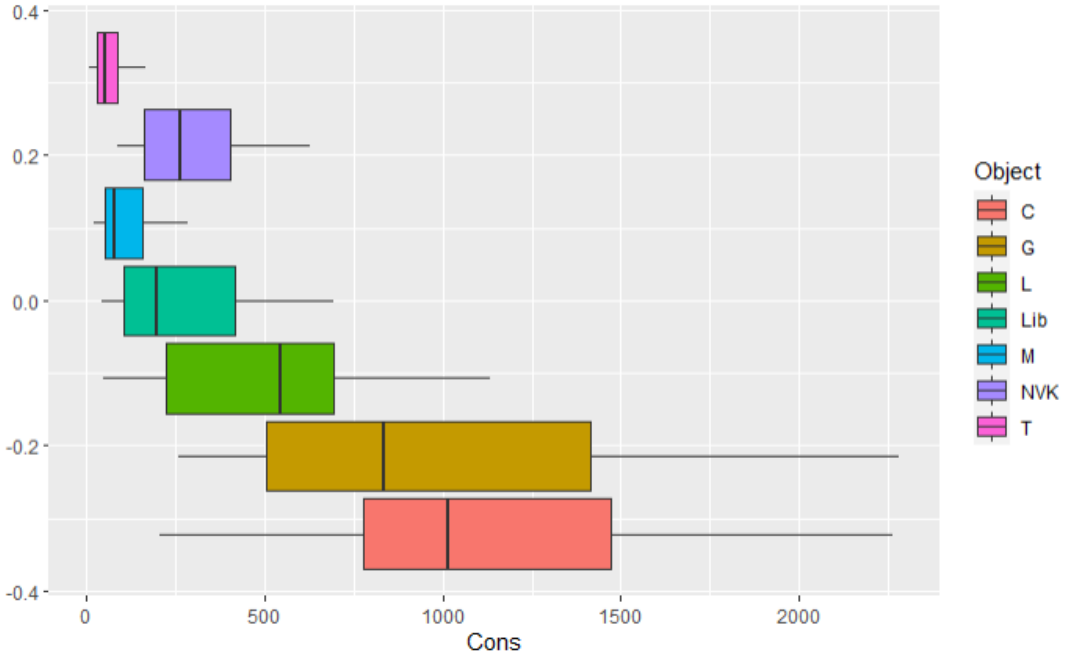


Рис. 3.3 – Описова статистика споживання електроенергії (*Cons*) за корпусами у 11’2017 – 10’2020 рр. на діаграмі *boxplot*

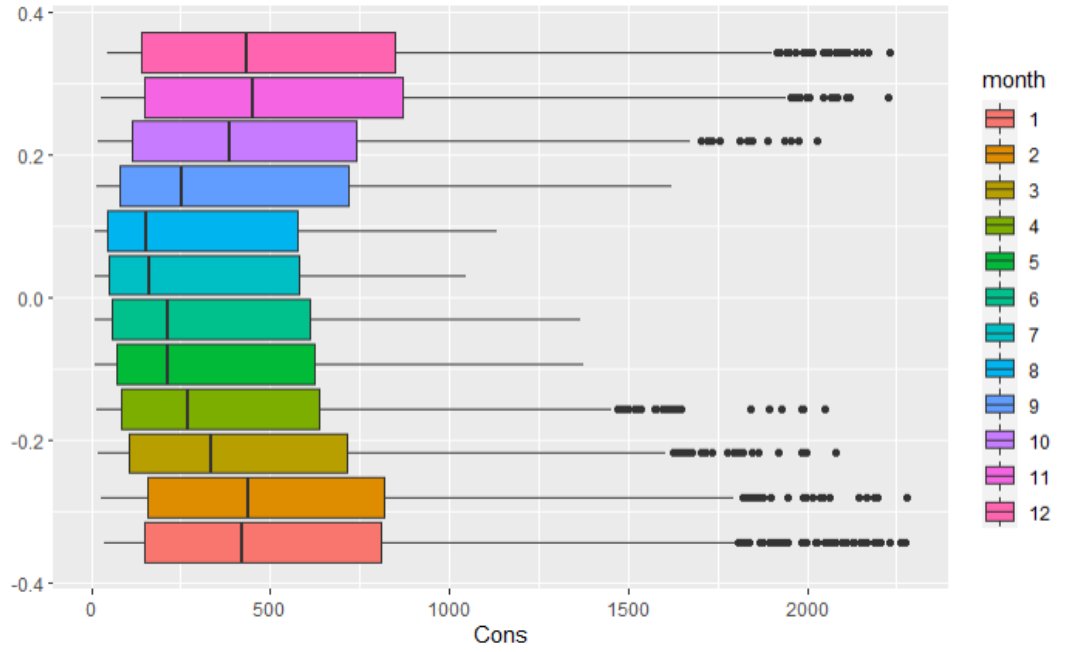


Рис. 3.4 – Описова статистика споживання електроенергії (*Cons*) за місяцями у 11’2017 – 10’2020 рр. на діаграмі *boxplot*

Наступним етапом аналізу є перевірка індивідуальних та панельних рядів на стаціонарність. Як показують результати висновувального аналізу, представлені у таблиці 3.2, загальний пуловий ряд споживання електроенергії (за повною вибіркою 2017-2020 рр., *panel*) є стаціонарним на 5% рівні значущості згідно розширеного тесту Дікі-Фулера та тесту Левіна- Ліна-Чу, проте є нестаціонарним згідно з тестом Хадрі, оскільки деякі індивідуальні ряди (за окремими корпусами) мають одиничні корені. Виявлено, що дані споживання за корпусом Ц (*panel_c$cons*) є нестаціонарними.

З точки зору подальшого аналізу, ми будемо брати прологарифмовані значення ряду споживання електроенергії без диференціювання для панельних моделей та, відповідно, для моделей часових рядів, що містять нестаціонарні дані – перші різниці прологарифмованих даних.

Таблиця 3.2 Тестування стаціонарності логарифмованих індивідуальних та панельних рядів споживання електроенергії (logcons) у 2017-2020 рр.

Вибірка / Тест на стаціонарність	**Augmented Dickey-Fuller Test** (На: пуловий або індивід. ряд є стаціонарний)		**Levin-Lin-Chu Unit-Root Test** (На: панельний ряд є стаціонарним)		**Hadri Test (Heteroskedasticity Consistent)** (На: як мін 1 ряд панелі має одиничні корені)	
Стат. / p-value	DF стат.	p-value	Z-стат.	p-value	Z-стат.	p-value
panel$logcons	-4.098	0.01*	-41.449	<2.2e-16	101.4	< 2.2e-16
panel_s1$logcons	-4.519	0.01*	-35.327	<2.2e-16	80.109	< 2.2e-16
panel_cg$logcons	-3.1628	0.09432	-21.686	<2.2e-16	63.032	<2.2e-16
panel_c$cons	-2.489	0.3713	-	-	-	-
panel_g$cons	-4.125	0.01*	-	-	-	-

* вказує на стаціонарність ряду на 5% рівні значущості.

З точки зору подальшого аналізу, ми будемо брати прологарифмовані значення ряду споживання електроенергії без диференціювання для панельних моделей та, відповідно, для моделей часових рядів, що містять нестаціонарні дані – перші різниці прологарифмованих даних.

Аналіз індивідуальних рядів споживання за корпусами Ц та Г у 2015-2020 рр. показує наявність місячної та тижневої сезонності (рис. 3.5).

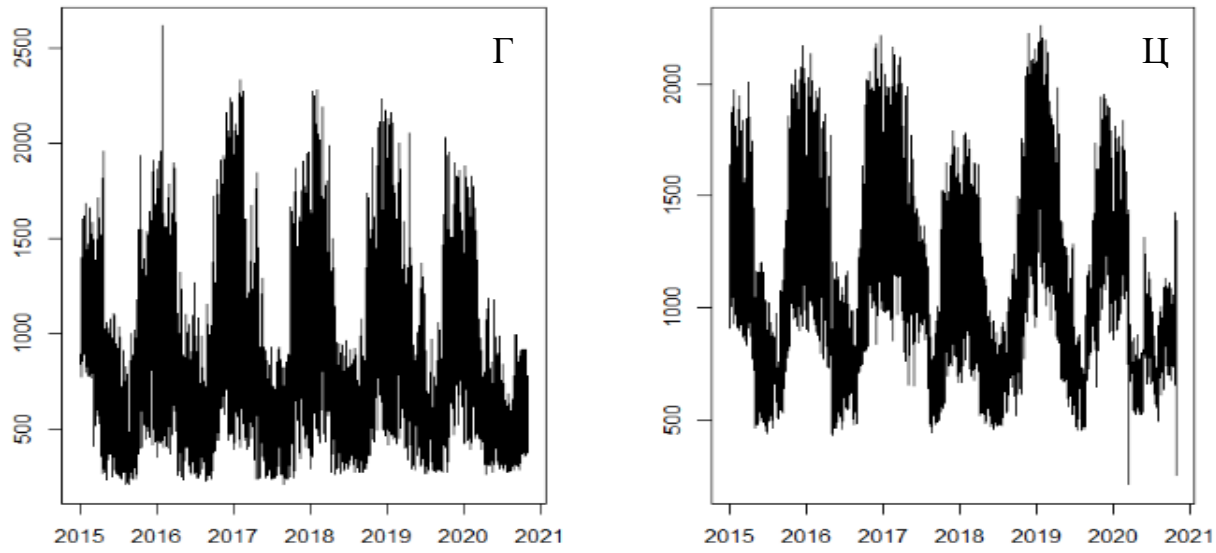


Рис. 3.5 – Динаміка електроспоживання у 2015 – 10’2020 рр. по корп. Г та Ц

Графічне зображення квантилів розподілу рядів споживання електроенергії по корпусам Г та Ц у 2015-2020 рр. показує значне відхилення від нормального розподілу (рис. 3.6).

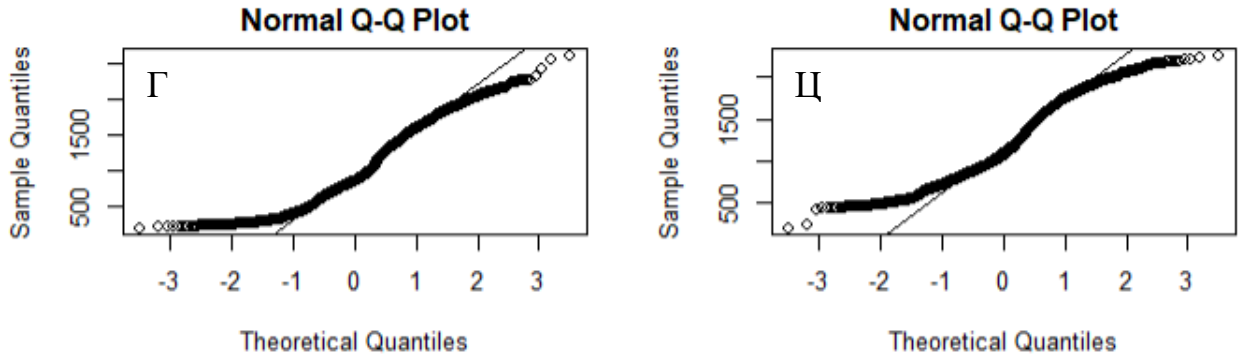


Рис. 3.6 – Графіки квантилів розподілу рядів споживання електроенергії по корп. Г та Ц

Класична мультиплікативна декомпозиція ряду [16] споживання електроенергії для корпусу Г (рис. 3.7) демонструє окрім сезонної компоненти ще і трендову та випадкові компоненти ряду. Тренд-циклічний компонент $\hat{T}_t$ обраховано як:

$$2 \times m - MA \tag{3.1}$$

Тут m – сезонний період ($m = 12$ для місячних даних та $m = 7$ для щоденних даних з тижневою сезонністю).

Ковзне середнє (MA) порядку m обраховується як:

$$\hat{T}_t = \frac{1}{m}\sum_{j=-k}^{k} y_{t+j}, \tag{3.2}$$

де $m = 2k + 1$; k – кількість періодів часу t.

Ряд без тренду обраховується як $y_t/\hat{T}_t$. Сезонний компонент отримуємо як середнє значення ряду буз тренду для кожного сезону (m).

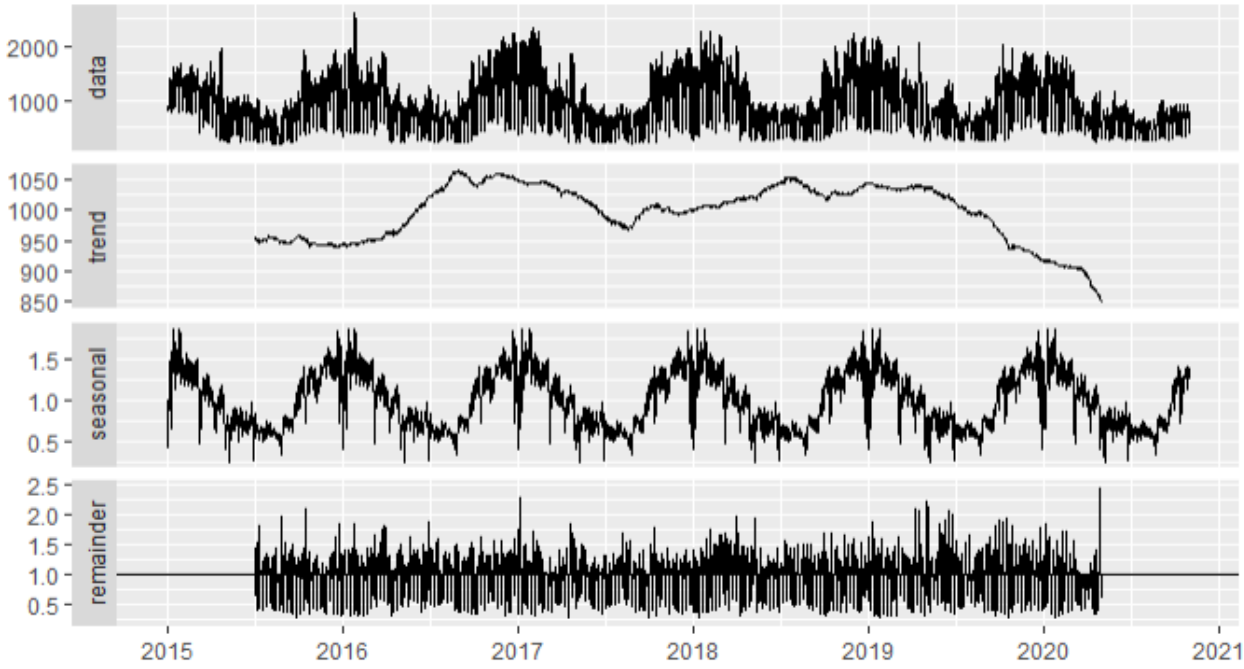


Рис. 3.7 – Мультиплікативна декомпозиція ряду споживання електроенергії корпусу Г

Як бачимо за візуалізаціями корпусу Г, ряд містить як сезонний компонент так і трендову динаміку, що пояснюється карантином 2020 р. Дані компоненти будуть враховані у відповідних моделях часових рядів.

3.2 Дослідження взаємозв'язків споживання електроенергії

Дослідження взаємозв'язків споживання електроенергії включає проведення кореляційного аналізу для виявлення лінійних стохастичних залежностей та причинно-наслідкових звязків за тестами Грейнджера та Думітреску-Хурліна [14]. Додатково проведено аналіз відмінностей у середніх рівнях електроспоживання, спричинених якісними параметрами (категорійними змінними) на підставі ANOVA-тесту (аналіз дисперсії) [16].

Результати кореляційного аналізу представлені на рис. 3.8.

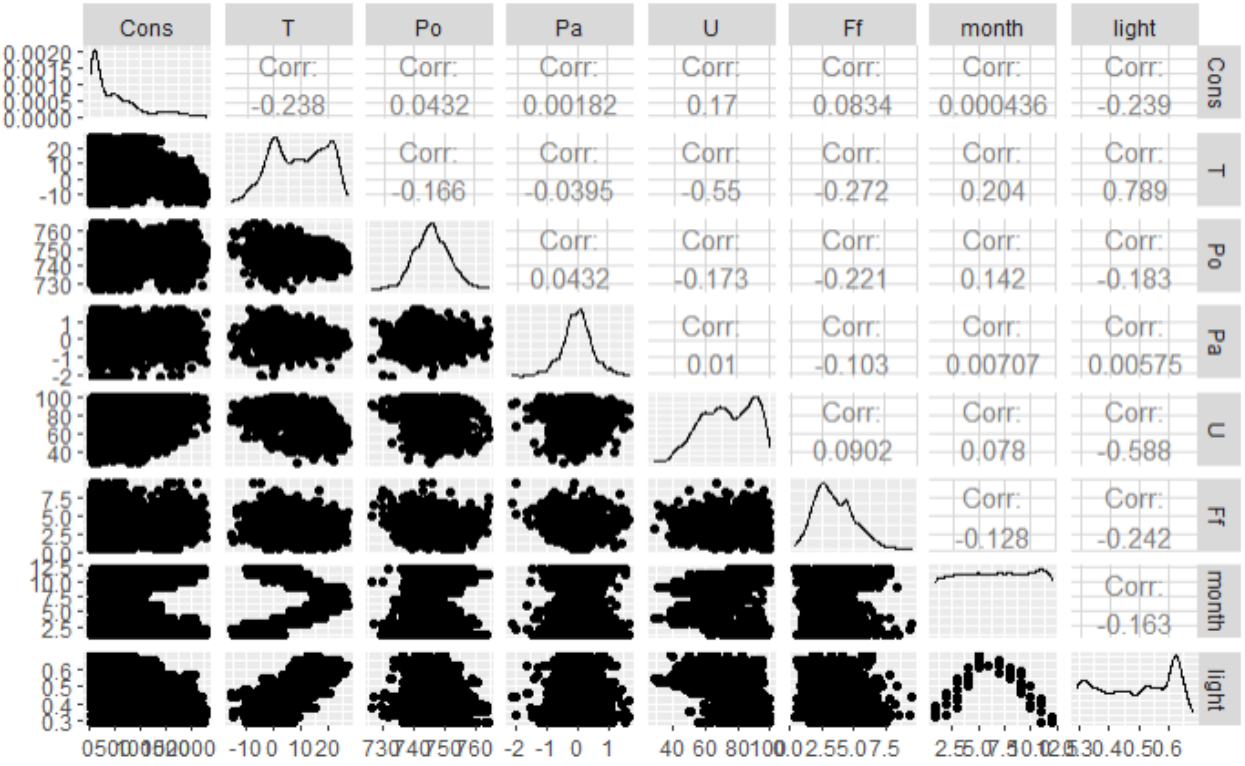


Рис. 3.8 – Діаграми розсіювання та кореляційна матриця для кількісних змінних загальної вибірки 2017-2020 рр.

Як бачимо, статистична сила зв'язку між змінною споживання електроенергії (*Cons*) та змінними погоди і календаря є невисокою. Найбільші значення від'ємної кореляції мають місце зі змінними освітлення (*light*) та температури (*T*). При цьому між змінними *light* та *T* спостерігається мультиколінеарність – статистично значущий кореляційний зв'язок.

Аналіз залежності споживання електроенергії по корпусу Г у 2015-2020 рр. з температурою представлено на рис. 3.9.

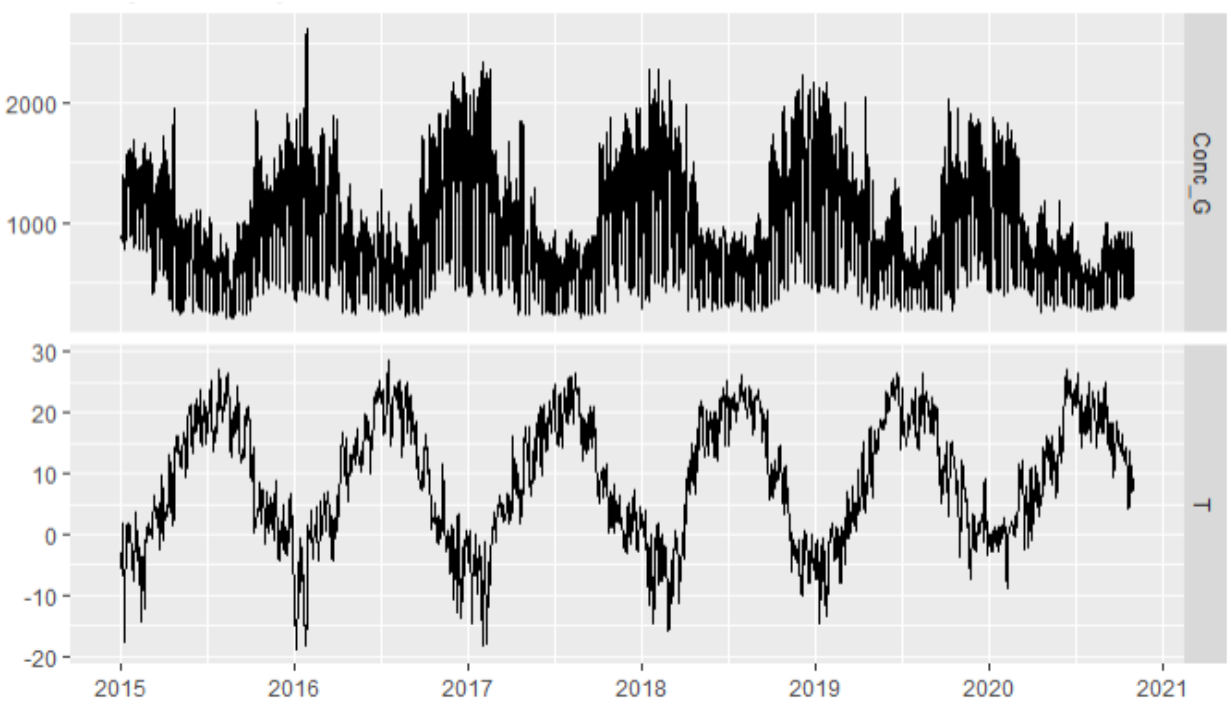


Рис. 3.9 – Динаміка рядів споживання електроенергії (корпус Г) та температури

Відмінності у середніх рівнях електроспоживання у розрізі категорійних змінних (корпуси, місяці, дні) представлені на рис. 3.10 -3.12.

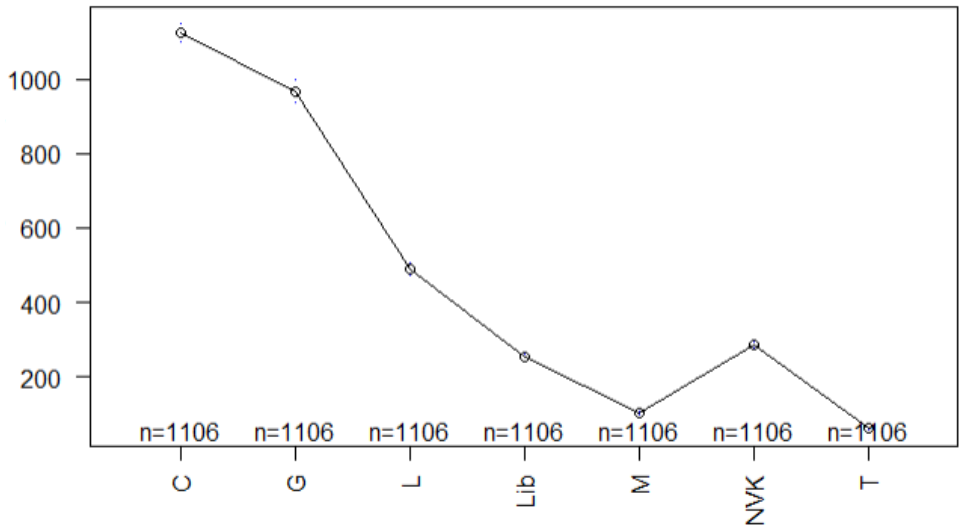


Рис. 3.10 – Середні значення електроспоживання у 11’2017 – 10’2020 рр. за корпусами

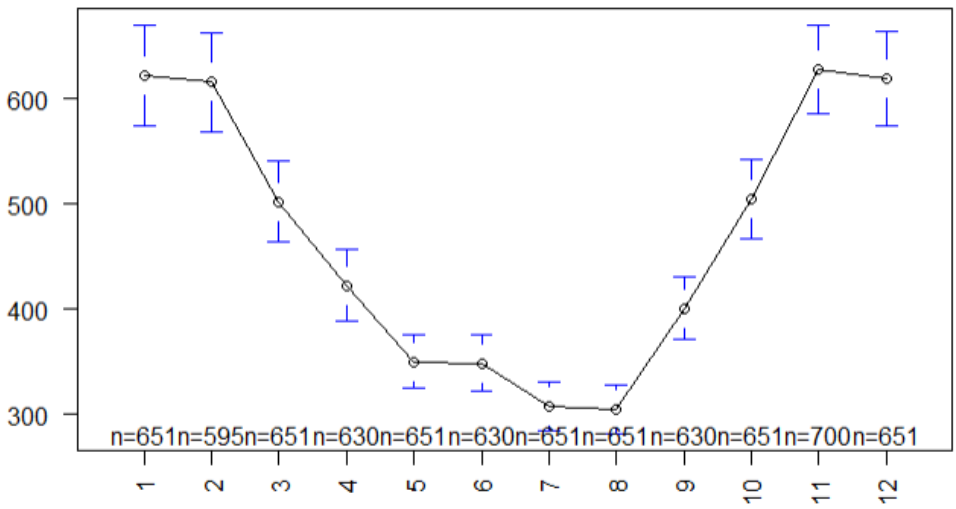


Рис. 3.11 – Середні значення електроспоживання у 11’2017 – 10’2020 рр. по місяцям

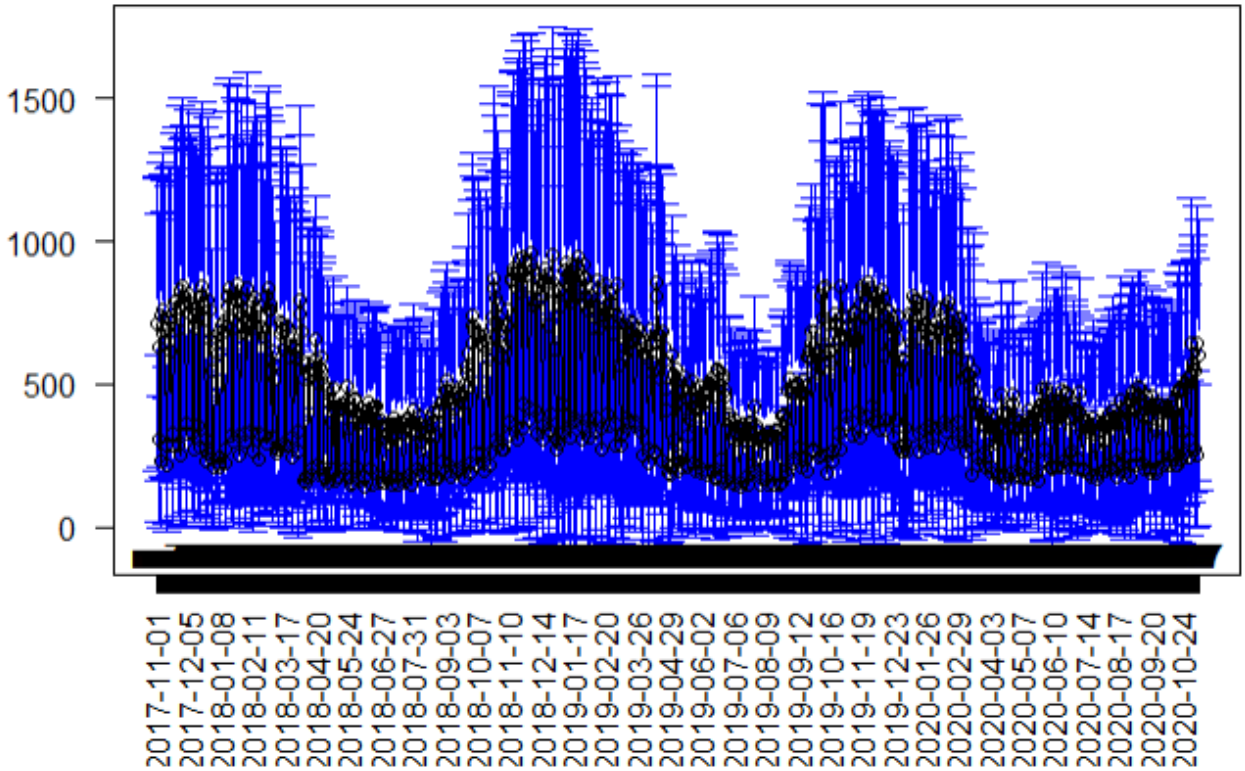


Рис. 3.12 – Середні значення електроспоживання у 11’2017 – 10’2020 рр. по дням

Рис 3.10 – 3.12 містять окрім середніх рівнів споживання електроенергії за категоріями ще довірчі інтервали середніх значень (CI), що обраховуються як:

$$CI = \bar{X} \pm t_{\alpha,d.f.} \frac{S}{\sqrt{n}}, \quad (3.3)$$

де $t_{\alpha,d.f.}$ – t-статистика для рівня значущості $\alpha = 0{,}05$ та ступенів свободи $df = n - 1$. $SE = t_{\alpha,d.f.} \frac{S}{\sqrt{n}}$ – стандартна похибка середнього $\bar{X}$, S – стандартне відхилення, n – кількість спостережень даних вибірки.

Статистичні значущість різних середніх значень споживання по вихідним та святковим дням перевірена ANOVA-тестом (аналіз дисперсії) [16]. F-статистика = 904.6 та 94.67, відповідно, що дає підстави відхилити нульову гіпотезу про рівність середніх на рівні значущості $\alpha = 0.001$.

Статистична відмінність у середніх значеннях споживання електроенергії найбільш значуща за об'єктами (корпусами), що підтверджується найбільшою F-статистикою = 3283. Різниці у середніх значеннях споживання між різними корпусами у 2017-2020 рр. представлена на рис. 3.13 результатами тесту порівняння середніх Тукі (*Tukey multiple comparisons of means*) [17]для рівня значущості $\alpha = 0.05$. Найістотніші відмінності позначені на рисунку жирним.

$Object	diff	lwr	upr	p adj
G-C	-0.2385469	-0.3197641	-0.1573297	0
L-C	-1.0107495	-1.0919667	-0.9295323	0
Lib-C	-1.6837879	-1.7650051	-1.6025707	0
M-C	**-2.5376071**	-2.6188243	-2.4563899	0
NVK-C	-1.4074486	-1.4886658	-1.3262314	0
T-C	**-3.0232510**	-3.1044682	-2.9420338	0
L-G	-0.7722026	-0.8534198	-0.6909854	0
Lib-G	-1.4452410	-1.5264582	-1.3640238	0
M-G	**-2.2990602**	-2.3802774	-2.2178430	0
NVK-G	-1.1689017	-1.2501189	-1.0876845	0
T-G	**-2.7847041**	-2.8659213	-2.7034869	0
Lib-L	-0.6730384	-0.7542556	-0.5918212	0
M-L	-1.5268576	-1.6080748	-1.4456404	0
NVK-L	-0.3966991	-0.4779163	-0.3154819	0
T-L	-2.0125015	-2.0937187	-1.9312843	0
M-Lib	-0.8538192	-0.9350364	-0.7726020	0
NVK-Lib	0.2763393	0.1951221	0.3575565	0
T-Lib	-1.3394631	-1.4206803	-1.2582459	0
NVK-M	1.1301585	1.0489413	1.2113757	0
T-M	-0.4856439	-0.5668611	-0.4044267	0
T-NVK	-1.6158024	-1.6970196	-1.5345853	0

Рис. 3.13 – Результати тесту порівняння середніх Тукі

Дослідження причинно-наслідкових зв’язків проведено для щоденних та агрегованих місячних у два етапи:

1) аналіз причинності за Грейнджером для пулових даних

2) аналіз причинності з урахуванням гетерогенності об’єктів панелі за тестом Думітреску-Хурліна.

Нульовою гіпотезою у обох випадках є:

1) H_0: x не спричиняє Granger y для всіх об’єктів (ряд розглядається як часовий ряд);

2) H_0: x не спричиняє Granger y для більшості об’єктів (ряд розглядається як панельний ряд, коефіцієнти тестової моделі можуть змінюватися для різних об’єктів).

Виявлено, що минулі значення (досліджувалися лаги 1-7 у тижневому інтервалі) таких змінних покращують прогнозування споживання щоденних показників електроенергії (порівняно із лише минулими значеннями змінної споживання електроенергії) при рівні значущості $\alpha = 0.001$:

- Температура (T) – причинність для лагів 1-4, 6-7;
- Рівень освітлення (light) – причинність для лагів 1-4, 7;
- Вологість (D) – для лагів 1-7.

Для щоденних даних з лагом $l = 30$ виявлено причинність між зазначеними вище змінними у обох напрямах: $x \sim y$ та $y \sim x$.

Аналіз агрегованих місячних даних через програму EViews [13] підтвердив причинність за Грейнджером на лагах 2-3 для споживання електроенергії від змінних *T, light, D, month, Holiday* (Додаток Б). Перевірка гіпотез за тестом Думітреску-Хурліна показала такі ж результати для усіх змінних, окрім *D*. Додаткове дослідження на місячних лагах $l = 6$ та $l = 9$ підтвердило причинність від змінних *Holiday, light, month* (Додаток Б).

3.3 Панельні лінійні регресійні моделі споживання електроенергії

Відповідно до алгоритму відбору оптимальної панельної моделі (див. рис.1.1), використовуючи методику покрокового регресійного аналізу, було спочатку оцінено чотири багатофакторні пулові МНК-моделі (OLS, OLS0, OLS1, OLS2) за всіма спостереженнями 7 корпусів. Покрокова процедура побудови багатофакторної регресії полягає в тому, що в кінцевій моделі мають залишитися лише значущі коефіцієнти, ймовірність нульової гіпотези (H_0: β_i=0) для яких невелика: *prob.* < 0.05 (зазвичай, для економічних розрахунків обирають рівень значущості α = 0.05, що відповідає 95% надійності). Коефіцієнт перетину у МНК-моделях представляє спільну для всіх компаній константу *c*.

Оцінки параметрів та метрик якості моделей зі значущими регресійними коефіцієнтами представлені на рис. 3.14.

```
%           & OLS      & OLS0     & OLS1     & OLS2    \\
(Intercept) & 6.11 ^*   & 6.00 ^*  & 5.68 ^*  & 7.15 ^* \\
            & (0.02)    & (0.03)   & (0.08)   & (0.06)  \\
T           & -0.03 ^*  & -0.03 ^* & -0.03 ^* &         \\
            & (0.00)    & (0.00)   & (0.00)   &         \\
Weekend1    & -0.93 ^*  & -0.93 ^* & -0.94 ^* & -0.93 ^*\\
            & (0.03)    & (0.03)   & (0.03)   & (0.03)  \\
Holiday1    & -0.97 ^*  & -0.96 ^* & -0.96 ^* & -0.99 ^*\\
            & (0.07)    & (0.07)   & (0.07)   & (0.07)  \\
month       &           & 0.02 ^*  & 0.01 ^*  & -0.01 ^*\\
            &           & (0.00)   & (0.00)   & (0.00)  \\
U           &           &          & 0.00 ^*  &         \\
            &           &          & (0.00)   &         \\
light       &           &          &          & -2.53 ^*\\
            &           &          &          & (0.10)   \\
 $N$         & 7742      & 7742     & 7742     & 7742     \\
$R^2$       & 0.19      & 0.19     & 0.19     & 0.19     \\
adj. $R^2$  & 0.19      & 0.19     & 0.19     & 0.19     \\
Resid. sd   & 1.10      & 1.10     & 1.10     & 1.10      \\
```

Рис. 3.14 – Результати МНК-оцінок пулових панельних моделей

Доцільність використання моделей з більшою кількістю коефіцієнтів було перевірено anova-тестом, який оцінює значущість додаткових змінних та їх додатковий внесок у пояснення залежної змінної. Хоча значення F-статистики свідчило на користь найбільш повної моделі, було обрано модель OLS0 замість OLS1 через проблему мультиколінеарності міжзмінними T та U.

Для перевірки необхідності включення у модель індивідуальних фіксованих ефектів крос-секційних груп (корпусів) та випадкових ефектів (збурень) нами було оцінено відповідні моделі та проведені необхідні тести згідно процедури, визначеної на рис. 1.1.

Регресійне рівняння панельної моделі з фіксованими ефектами має такий вигляд (усі змінні значущі при $\alpha = 0.0001$):

$$log(cons)_{it} = 7.41 - 0.031T_{it} - 0.933Weekend1 - 0.959Holiday1 + 0.019month -$$
$$-0.239\text{factor(Object)G} - 1.011\text{factor(Object)L} - 1.408\text{factor(Object)NVK} -$$
$$-1.684\text{factor(Object)Lib} - 2.538\text{factor(Object)M} - 3.023\text{factor(Object)T} \quad (3.4)$$

Коефіцієнти пояснювальних змінних моделі відповідають оцінкам моделей (рис. 3.14). Коефіцієнти фіктивних змінних об'єктів (*factor(Object)*) показують зміщення по відношенню до базового об'єкта, виключеного з моделі – *factor(Object)C* – показників споживання Центрального корпусу.

Показники адекватності моделі (3.4):

- Residual standard error: 0.3711 on 7731 degrees of freedom
- Multiple R-squared: 0.9074, Adjusted R-squared: 0.9073
- F-statistic: 7579 on 10 and 7731 DF, p-value: < 2.2e-16

Графіки залишків моделі (3.4), представлені на рис. 3.15, в цілому підтверджують адекватність моделі – гомоскедастичність та нормальний розподіл залишків (за винятком незначної кількості аномалій), відсутність серійної кореляції. Прогнозну якість підтверджує також діаграма розсіювання значень, оцінених за моделлю (3.4), та фактичних значень споживання електроенергії (рис. 3.16). Рис. 3.17 показує усі лінії регресії за (3.4).

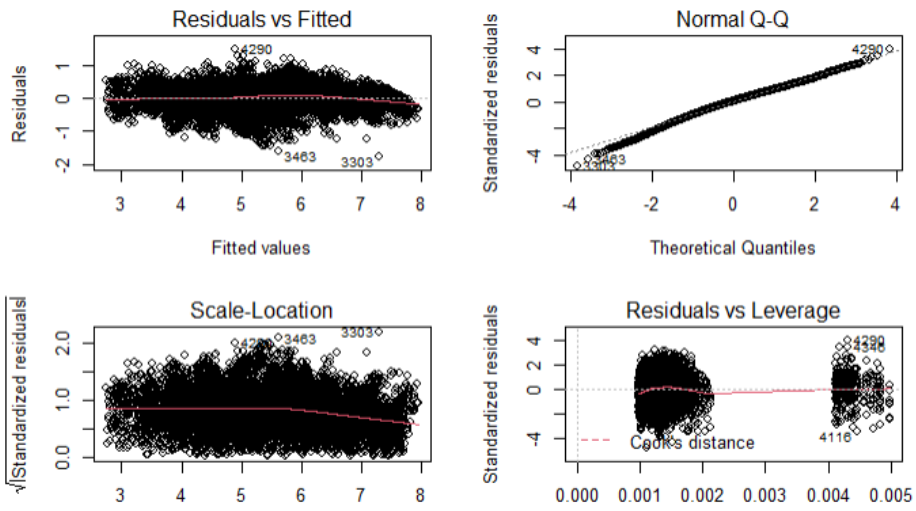


Рис. 3.15 – Графіки залишків моделі (3.4)

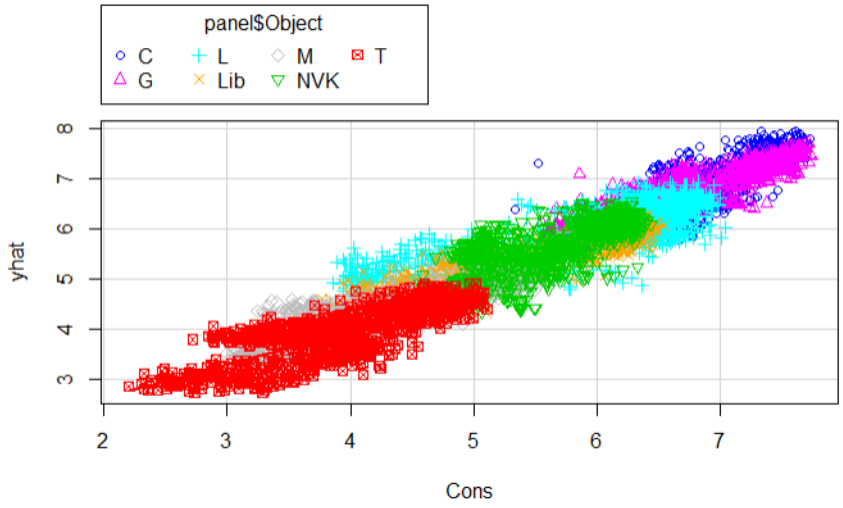


Рис. 3.16 – Значення Cons: прогнозні проти фактичних за моделлю (3.4)

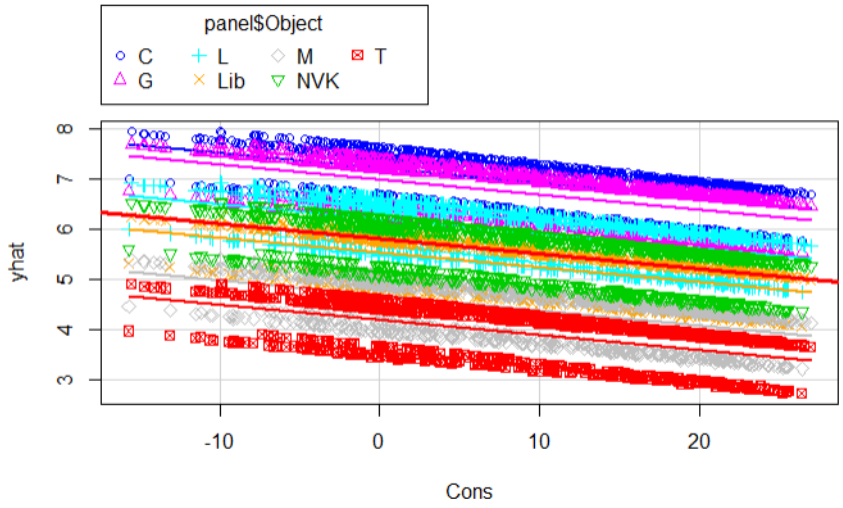


Рис. 3.17 – Прогнозні значення Cons за моделлю (3.4): середні за об’єктами

На рис. 3.18 представлена діаграма розсіювання споживання електроенергії та температури повітря, що показує апроксимацію прогнозних та фактичних даних за моделлю (3.4).

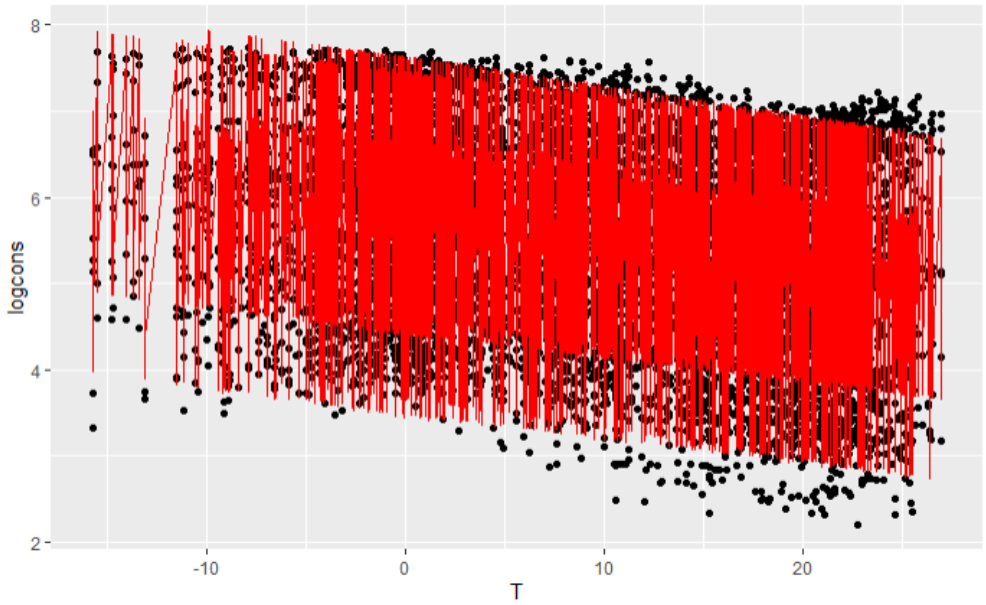


Рис. 3.18 – Діаграма розсіювання прогнозних і фактичних значень Cons та T за (3.4)

Рівняння панельної моделі з фіксованими ефектами (3.4), оцінене з використанням Within Estimator (1.8) алгоритму бібліотеки R plm [19], має вигляд:

$$log(cons)_{it} = -0.031T_{it} - 0.933Weekend1 - 0.959Holiday1 + 0.019month \qquad (3.5)$$

При цьому оцінки індивідуальних ефектів об’єктів за (3.5) є такими:

```
    Estimate Std. Error t-value  Pr(>|t|)
C   7.410431   0.014234  520.62 < 2.2e-16 ***
G   7.171884   0.014234  503.86 < 2.2e-16 ***
L   6.399682   0.014234  449.61 < 2.2e-16 ***
Lib 5.726643   0.014234  402.33 < 2.2e-16 ***
M   4.872824   0.014234  342.34 < 2.2e-16 ***
NVK 6.002982   0.014234  421.74 < 2.2e-16 ***
T   4.387180   0.014234  308.22 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Оцінки різниці індивідуальних ефектів об’єктів від загального середнього значення споживання за (3.5) є такими:

```
      Estimate Std. Error   t-value Pr(>|t|)
C    1.4144844  0.0142339   99.3746   <2e-16 ***
G    1.1759375  0.0142339   82.6155   <2e-16 ***
L    0.4037349  0.0142339   28.3644   <2e-16 ***
Lib -0.2693035  0.0142339  -18.9199   <2e-16 ***
M   -1.1231227  0.0142339  -78.9050   <2e-16 ***
NVK  0.0070358  0.0142339    0.4943   0.6211
T   -1.6087666  0.0142339 -113.0239   <2e-16 ***
```

Показники адекватності моделі (3.5):

- Total Sum of Squares: 3244.2
- Residual Sum of Squares: 1065
- R-Squared: 0.67173
- Adj. R-Squared: 0.67131
- F-statistic: 3954.97 on 4 and 7731 DF, p-value: < 2.22e-16

Відповідно до процедури відбору оптимальної панельної моделі (див. рис.1.1) було додатково оцінено модель з випадковими ефектами (1.6 – 1.7), коефіцієнти якої відрізнялися від моделі з фіксованими ефектами (3.5) лише значеннями константи:

$$log(cons)_{it} = 5.996 - 0.031T_{it} - 0.933Weekend1 - 0.959Holiday1 + 0.019month \quad (3.6)$$

Відповідно до тесту Хаусмана [19], ми не можемо прийняти нульову гіпотезу щодо ендогенності предикторів (відсутність кореляції між предикторами та похибками моделі), і приймаємо альтернативну гіпотезу, згідно з якою оптимальною є модель з фіксованими ефектами.

Подальший аналіз панельної моделі (3.5) включав [20], [21]:

1. Перевірку крос-секційної залежності панельних даних у моделі за допомогою тесту Pesaran CD test – z = 25.702, p-value < 2.2e-16 – приймаємо альтернативну гіпотезу щодо крос-секційної залежності даних.

2. Перевірку значущості часових ефектів з Lagrange Multiplier Test - time effects (Breusch-Pagan) for balanced panels – chisq = 351.83, df = 1, p-value

< 2.2e-16 – приймаємо альтернативну гіпотезу щодо необхідності включення часових ефектів до моделі.

3. Перевірка серійної кореляції залишків панельних моделей з Breusch-Godfrey/Wooldridge test for serial correlation – chisq = 5367.8, df = 1106, p-value < 2.2e-16 – приймаємо альтернативну гіпотезу щодо наявності серійної кореляції залишків.

4. Перевірка гомоскедастичності залишків панельних моделей з Breusch-Pagan test – BP = 95.723, df = 4, p-value < 2.2e-16 – приймаємо альтернативну гіпотезу щодо гетероскедастичності залишків.

Одним з варіантів вирішення проблем, виявлених тестами 1-4 є побудова моделі з фіксованими ефектами з включенням лагів змінної споживання електроенергії та оцінка коефіцієнтів за робастними методами, що є стійкими до наявної гетероскедастичності та автокореляції залишків [21]:

$$log(cons)_{it} = 0.189\log(cons)_{it-1} + 0.099\log(cons)_{it-3} + 0.45\log(cons)_{it-7} - \\ -0.73\log(cons)_{it-30} - 0.01T_{it} - 0.505Weekend1 - 0.85Holiday1 + 0.004month \quad (3.7)$$

Дана модель має кращий за модель (3.5) коефіцієнт детермінації R^2=0.8195, скорегований на ступені свободи (df=7517) R^2=0.8192.

Робастні оцінки коефіцієнтів відповідають оцінкам (3.7), при цьому залишаючись значущими на рівні значущості $\alpha = 0.001$. Порівняння оцінок коефіцієнтів оригінальної (3.5) та робастної моделі (3.7) наведено нижче.

Оригінальна модель 3.5 - t test of coefficients:

```
                   Estimate  Std. Error  t value  Pr(>|t|)
lag(logcons, 1)   0.18895811  0.00663084  28.4969 < 2.2e-16 ***
lag(logcons, 3)   0.09931343  0.00608588  16.3187 < 2.2e-16 ***
lag(logcons, 7)   0.45007476  0.00812510  55.3931 < 2.2e-16 ***
lag(logcons, 30) -0.07296043  0.00626690 -11.6422 < 2.2e-16 ***
T                -0.01027281  0.00044229 -23.2264 < 2.2e-16 ***
Weekend1         -0.50468620  0.01032270 -48.8909 < 2.2e-16 ***
Holiday1         -0.85043209  0.01677580 -50.6940 < 2.2e-16 ***
month             0.00375608  0.00097804   3.8404 0.0001238 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Робастна модель 3.7 - t test of coefficients:

```
                   Estimate  Std. Error  t value  Pr(>|t|)
lag(logcons, 1)   0.18895811  0.01759575  10.7388 < 2.2e-16 ***
lag(logcons, 3)   0.09931343  0.01707851   5.8151 6.307e-09 ***
lag(logcons, 7)   0.45007476  0.03199564  14.0668 < 2.2e-16 ***
lag(logcons, 30) -0.07296043  0.00656805 -11.1084 < 2.2e-16 ***
T                -0.01027281  0.00170268  -6.0333 1.682e-09 ***
Weekend1         -0.50468620  0.10069076  -5.0122 5.503e-07 ***
Holiday1         -0.85043209  0.12673170  -6.7105 2.081e-11 ***
month             0.00375608  0.00078411   4.7902 1.698e-06 ***
---
```

Приклад прогнозу споживання електроенергії (точковий та прогнозні інтервали) за заданих вхідних параметрів згідно з моделлю (3.4) наведено у таблиці 3.4.

Таблиця 3.4 – Прогнозні оцінки споживання електроенергії за лінійною регресійною моделлю з фіксованими ефектами

Вхідні параметри	Довірчі інтервали	Прогнозні інтервали	Середні фактичні значення за даними 2019 р.
T=2, Weekend='0', Holiday='0', month=12, Object='G'	fit = 1528.876 lower = 1487.422 upper = 1571.485	fit = 1528.876 lower = 738.202 upper = 3166.425	fit = 1552.08* * обраховано як середнє значення за T ~ 2
T=2, Weekend='1', Holiday='0', month=12, Object='G'	fit = 601.3323 lower = 583.6025 upper = 619.6007	fit = 601.3323 lower = 290.3191 upper = 1245.528	fit = 627.96* * обраховано як середнє значення за T ~ 2

Якщо довірчі інтервали прогнозних оцінок згідно з табл. 3.4 є доволі прийнятними, то прогнозні інтервали, які враховують випадкові непередбачувані похибки є високими. Це говорить про необхідність огляду специфічних моделей для часових рядів – авторегресійних моделей розподіленого лагу (ARDL) для панельних рядів та різновиду моделей часових рядів для індивідуальних рядів.

Повне представлення програмної реалізації в інтегрованому середовищі розробки IDE RStudio представлено у Додатку В.

3.4 Панельні авторегресійні моделі розподіленого лагу

Надалі ми додатково використали підхід панельних авторегресійних моделей розподіленого лагу (ARDL) для оцінки короткострокових та довгострокових зв'язків між змінними. Даний метод моделювання може бути застосований як для часових, так і для панельних рядів. Оскільки за щоденними даними важко оцінити довгострокові зв'язки, ми усереднили дані і створили вибірку з місячними спостереженнями 11'2017 – 10'2020 рр.

Оцінювання моделей ARDL проведено у статистичному пакеті EViews 10. Вибір оптимальної кількості лагів для залежної та пояснювальних змінних моделі було здійснено за критерієм мінімізації інформаційного критерію AIC (табл. 3.5).

Таблиця 3.5 – Вибір порядку лагу для моделі ARDL

Model Selection Criteria Table
Dependent Variable: LOGCONS
Date: 12/20/20 Time: 23:34
Sample: 2017M11 2020M11
Included observations: 259

Model	LogL	AIC*	BIC	HQ	Specification
16	164.471699	**-0.367720**	1.450353	0.365572	**ARDL(4, 4, 4, 4)**
12	152.402816	-0.323834	1.389924	0.367384	ARDL(3, 4, 4, 4)
8	139.155795	-0.269747	1.339695	0.379397	ARDL(2, 4, 4, 4)
4	130.979073	-0.259559	1.245567	0.347511	ARDL(1, 4, 4, 4)
10	102.136542	-0.252264	0.835599	0.186509	ARDL(3, 2, 2, 2)
5	73.931336	-0.250488	0.420113	0.019989	ARDL(2, 1, 1, 1)
6	93.251361	-0.235943	0.747605	0.160757	ARDL(2, 2, 2, 2)
9	78.887792	-0.232795	0.542122	0.079756	ARDL(3, 1, 1, 1)
11	118.461197	-0.211785	1.189025	0.353210	ARDL(3, 3, 3, 3)
14	104.034423	-0.208090	0.984089	0.272757	ARDL(4, 2, 2, 2)
7	108.816047	-0.188884	1.107611	0.334038	ARDL(2, 3, 3, 3)
13	80.391013	-0.185204	0.694029	0.169421	ARDL(4, 1, 1, 1)
15	122.017016	-0.181966	1.323161	0.425104	ARDL(4, 3, 3, 3)
3	97.607516	-0.152446	1.039733	0.328401	ARDL(1, 3, 3, 3)
2	75.048487	-0.138948	0.740284	0.215677	ARDL(1, 2, 2, 2)
1	49.148846	-0.096527	0.469758	0.131876	ARDL(1, 1, 1, 1)

Для визначеного порядку лагів залежної та пояснювальних змінних отримано оцінки коефіцієнтів моделі (табл. 3.6). Якщо загальна початкова

кількість спостережень становила 259, то після диференціювання (D) кількість спостережень для оцінювання параметрів становила 231.

Таблиця 3.6 – Оцінки коефіцієнтів моделі ARDL

Dependent Variable: D(LOGCONS)
Method: ARDL

Date: 12/20/20 Time: 23:32
Sample: 2018M03 2020M11
Included observations: 231
Maximum dependent lags: 4 (Automatic selection)
Model selection method: Akaike info criterion (AIC)
Dynamic regressors (4 lags, automatic): T WEEKEND HOLIDAY
Fixed regressors: C
Number of models evalulated: 16
Selected Model: ARDL(4, 4, 4, 4)
Note: final equation sample is larger than selection sample

Variable	Coefficient	Std. Error	t-Statistic	Prob.*
		Long Run Equation		
T	0.273787	0.393141	0.696408	0.4874
WEEKEND	-194.9999	242.5414	-0.803986	0.4228
HOLIDAY	-230.6419	297.9819	-0.774013	0.4403
		Short Run Equation		
COINTEQ01	-0.053488	0.019587	-2.730718	0.0072
D(LOGCONS(-1))	-0.172812	0.116239	-1.486703	0.1394
D(LOGCONS(-2))	-0.404563	0.072901	-5.549497	0.0000
D(LOGCONS(-3))	-0.279064	0.063001	-4.429531	0.0000
D(T)	-0.014629	0.004095	-3.572800	0.0005
D(T(-1))	-0.016457	0.003123	-5.269529	0.0000
D(T(-2))	-0.004744	0.003712	-1.278018	0.2034
D(T(-3))	-0.024684	0.008187	-3.014841	0.0031
D(WEEKEND)	6.874357	2.531745	2.715264	0.0075
D(WEEKEND(-1))	4.532238	1.902169	2.382669	0.0186
D(HOLIDAY)	9.474494	3.376193	2.806265	0.0057
D(HOLIDAY(-1))	7.568197	2.515376	3.008774	0.0031
C	3.549158	1.310271	2.708720	0.0076

Mean dependent var	-0.005338	S.D. dependent var	0.286298
S.E. of regression	0.155533	Akaike info criterion	-0.327967
Sum squared resid	3.314107	Schwarz criterion	1.347450
Log likelihood	164.4717	Hannan-Quinn criter.	0.345650

Оцінки короткострокових коефіцієнтів моделі ARDL(4, 4, 4, 4) наведені у Додатку В. Якщо більшість коефіцієнтів короткострокової динаміки (*Short Run Equation* у табл. 3.6) є значущими при рівні значущості $\alpha = 0.05$, то коефіцієнти довгострокової динаміки (*Long Run Equation* у табл. 3.6) є

незначущими. В результаті оцінки спрощеної моделі ARDL, за виключенням змінних Weekend та Holiday, було отримано значущі коефіцієнти довгострокового рівняння (табл. 3.7):

Таблиця 3.7 – Оцінки коефіцієнтів спрощеної моделі ARDL

Dependent Variable: D(LOGCONS)
Method: ARDL
Date: 12/20/20 Time: 23:01
Sample: 2018M01 2020M11
Included observations: 245
Maximum dependent lags: 4 (Automatic selection)
Model selection method: Akaike info criterion (AIC)
Dynamic regressors (4 lags, automatic): T
Fixed regressors: C
Number of models evalulated: 16
Selected Model: ARDL(2, 1)
Note: final equation sample is larger than selection sample

Variable	Coefficient	Std. Error	t-Statistic	Prob.*
Long Run Equation				
T	-0.027812	0.002748	-10.12145	0.0000
Short Run Equation				
COINTEQ01	-0.536813	0.100130	-5.361142	0.0000
D(LOGCONS(-1))	0.335348	0.106717	3.142407	0.0019
D(T)	-0.005703	0.001859	-3.067554	0.0024
C	3.085752	0.591997	5.212443	0.0000
Mean dependent var	-0.005888	S.D. dependent var		0.279399
S.E. of regression	0.198871	Akaike info criterion		-0.325647
Sum squared resid	9.096428	Schwarz criterion		0.072607
Log likelihood	71.17134	Hannan-Quinn criter.		-0.165525

Відповідно до результатів табл. 3.7 можемо зробити висновки, що між споживанням електроенергії та температурою повітря існує значущий довгостроковий рівноважний зв'язок, про що свідчить коефіцієнт коінтеграційного рівняння COINTEQ01=-0.536813. При цьому, швидкість пристосування змінних до довгострокового рівноважного значення є доволі високою: 10% підвищення температури приводить до 3% зменшення споживання електроенергії.

Графічне зображення апроксимації фактичних та прогнозних даних наведено на рис. 3.19.

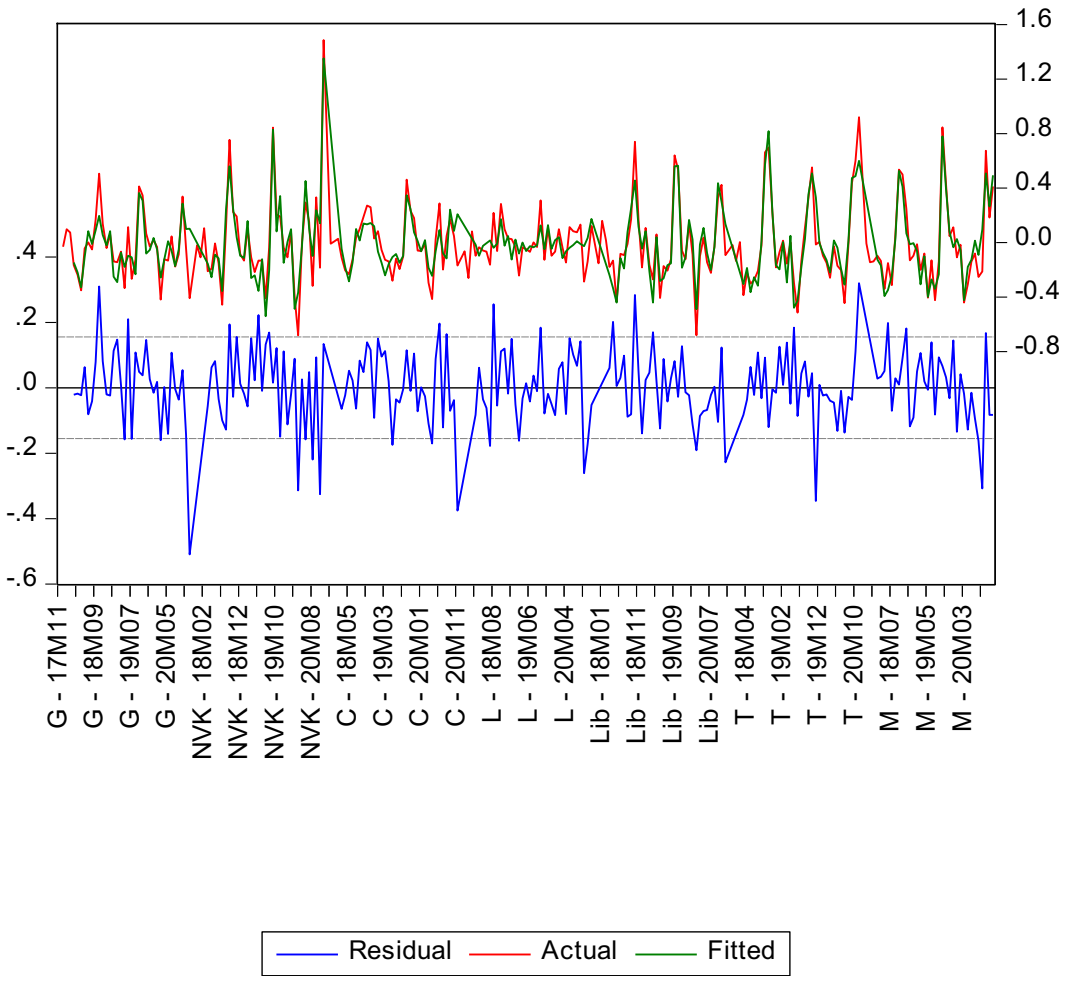
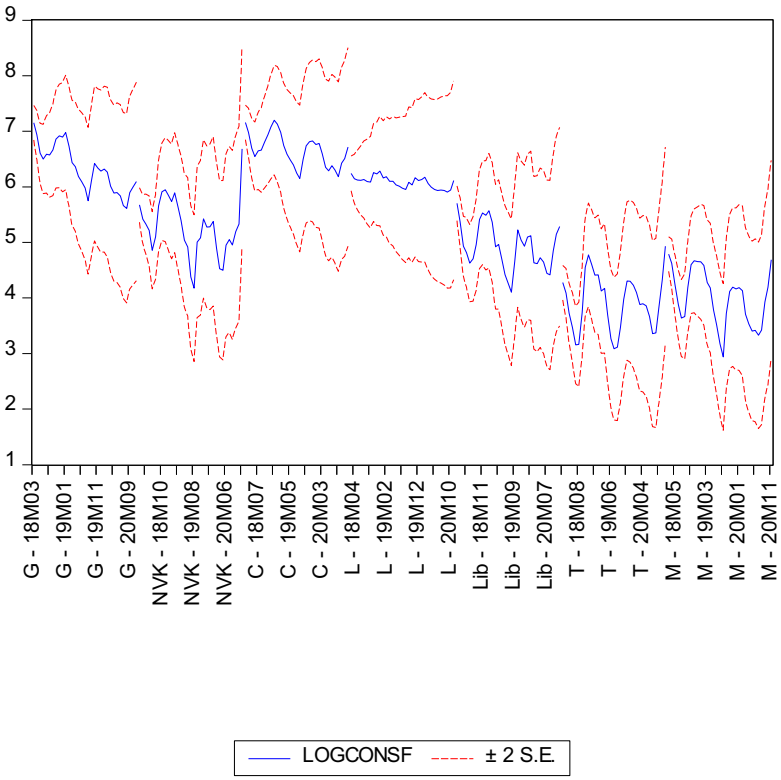


Рис. 3.19 – Фактичні, прогнозні дані та залишки моделі ARDL(4, 4, 4, 4)

На рис. 3.20 показані прогнозні інтервали за моделлю ARDL(4, 4, 4, 4) у розрізі корпусів, а також метрики прогнозної якості на тренувальній вибірці.

Як бачимо, показник відсоткової середньої похибки (MAPE) становить 6%. Відповідна метрика спрощеної моделі (табл. 3.7), що включає тільки температуру як пояснювальну змінну становить 9%.

Отримані результати показують гарну перспективу використання даного методу і можливість покращення результатів шляхом включення дієвих пояснювальних змінних, що відображують різницю у споживанні електроенергії за корпусами (кількість людей, оргтехніки тощо).



Forecast: LOGCONSF	
Actual: LOGCONS	
Forecast sample: 2017M11 2020M11	
Adjusted sample: 2018M03 2020M11	
Included observations: 231	
Root Mean Squared Error	0.415617
Mean Absolute Error	0.329520
Mean Abs. Percent Error	6.028557
Theil Inequality Coefficient	0.037309
Bias Proportion	0.507233
Variance Proportion	0.002008
Covariance Proportion	0.490758
Theil U2 Coefficient	1.263343
Symmetric MAPE	6.315253

Рис. 3.20 – Прогнозні інтервали за моделлю ARDL(4, 4, 4, 4) у розрізі корпусів

Інший метод оцінювання панельних моделей, який вирішує проблему ендогенності регресорів є узагальнений метод моментів (Generalized Method of Mometns / Dynamic Panel Data. Специфіка даного методу полягає у виборі інструментальної змінної (зазвичай, 2-ий або інший лаг залежної змінної) та оцінювання моделі за допомогою двокрокового методу найменших квадратів. На жаль, для наших емпіричних даних зазначений метод неможливо було використати через обмежену кількість спостережень, яка не дозволила обрахувати оцінки коефіцієнтів інструментальної змінної.

ВИСНОВКИ

В ході дослідження було проаналізовано сучасні тенденції у галузі електроенергетики, розглянуто підходи до моделювання та прогнозування та споживання електроенергії на рівні резидентного сектору та організацій, установ, підприємств.

Досліджено підходи лінійних регресійних та динамічних панельних моделей як дієвий механізм отримання статистично значущих результатів в умовах обмеженості та неоднорідності даних споживання електроенергії.

Розроблено алгоритм моделювання та прогнозування показників електроспоживання та визначення причинно-наслідкових зв’язків у короткостроковій та довгостроковій перспективі. Представлений алгоритм оцінено на емпіричних даних споживання електроенергії корпусами Центрального кампусу СумДУ. Підтверджено короткострокові зв’язки між показниками електроспоживання та температурою, рівнем освітлення, вологості, а також календарними ефектами (вихідні та святкові дні, місяць). Встановлено довгострокові рівноважні зв’язки між показниками електроспоживання та температурою. У той же час зроблено висновок щодо необхідності пошуку додаткових даних, які б пояснювали відмінності у споживанні електроенергії у різних корпусах. Це можуть бути кількість людей, що відвідують приміщення, кількість працюючої оргтехніки, приладів та пристроїв та їх клас енергоефективності, показники завантаження аудиторій згідно з розкладом навчальних занять тощо.

У результаті порівняльного аналізу лінійних регресійних панельних моделей та динамічних панельних моделей розподіленого лагу підтверджено доцільність застосування останніх за умови зменшення частоти даних та наявності достатньо великого розміру вибірки. Запропоновано застосовувати панельні моделі як доповнення до прогнозної аналітики часових рядів для покращення управлінських рішень щодо енергоспоживання і ефективності.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. GDP per unit of energy use (PPP $ per kg of oil equivalent). Data. *The World Bank.* URL: https://data.worldbank.org/indicator/EG.GDP.PUSE.KO.PP (Last accessed: 03.11.2020).
2. Data and statistics. *International Energy Agency*. URL: https://www.iea.org/data-and-statistics?country=UKRAINE&fuel=Electricity%20and%20heat&indicator=ElecConsPerCapita (Last accessed: 03.11.2020).
3. Scopus. URL: http://www.scopus.com/ (Last accessed: 03.11.2020).
4. Kalimoldayev M., Drozdenko A., Koplyk I., Marinich T., Abdildayeva A., Zhukabayeva T. Analysis of modern approaches for the prediction of electric energy consumption. *Open Engineering.* 2020. Vol. 10, Issue 1. P. 350–361. DOI: https://doi.org/10.1515/eng-2020-0028 (Last accessed: 03.11.2020).
5. Sotnyk M., Marynych T., Drozdenko A., Leontiev P., Telizhenko O. Monitoring and forecasting systems for electricity consumption in educational institutions. *Theoretical aspects of modern engineering* : collective monograph / L. Hnes et al. ; International Science Group. Boston : Primedia eLaunch, 2020. P. 139–158. URL: https://essuir.sumdu.edu.ua/bitstream-download/123456789/80932/1/Sotnyk_Marynych_Monograph-USA-Technical-isg-konf.pdf (Last accessed: 03.11.2020).
6. АСКОЕ. *IKNET.* URL: https://iknet.com.ua/uk/askue/ (дата звернення: 20.11.2020).
7. Swan L. G., Ugursal V. I. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews.* 2009. Vol. 13, Issue 8. P. 1819–1835. DOI: 10.1016/j.rser.2008.09.033 (Last accessed: 03.11.2020).
8. Ghysels E., Kvedaras V., Zemlys V. Mixed Frequency Data Sampling Regression Models: The R Package midasr. *Journal of Statistical Software.*

2016. Vol. 72, Issue 4. DOI: 10.18637/jss.v072.i04 (Last accessed: 03.11.2020).

9. Hanck C., Arnold M., Gerber A., Schmelzer M. Introduction to Econometrics with R. Essen : University of Duisburg-Essen, 2020. URL: https://www.econometrics-with-r.org/index.html (Last accessed: 03.11.2020). License: CC BY-NC-SA 4.0.
10. Pesaran M. H. A simple panel unit root test in the presence of cross section dependence. *Journal of Applied Econometrics.* 2007. Vol. 22 , Issue 2, P. 265–312. DOI: 10.1002/jae.951 (Last accessed: 03.11.2020).
11. Le H. P., Sarcodie S. A. Dynamic linkage between renewable and conventional energy use, environmental quality and economic growth: Evidence from Emerging Market and Developing Economies. *Energy Reports.* 2020. Vol. 6. P. 965–973. DOI: https://doi.org/10.1016/j.egyr.2020.04.020 (Last accessed: 03.11.2020).
12. Amaluddin A. The Dynamic Link of Electricity Consumption, Internet Access and Economic Growth in 33 Provinces of Indonesia. *International Journal of Energy Economics and Policy.* 2020. Vol. 10, Issue 4. P. 309–317. DOI: https://doi.org/10.32479/ijeep.9249 (Last accessed: 20.11.2020).
13. EViews : website. URL: http://www.eviews.com/ (Last accessed: 20.11.2020).
14. Lopez L., Weber S. Testing for Granger causality in panel data : IRENE Working paper 17-03. University of Neuchatel. Institute of Economic Research, 2017. URL: https://www.unine.ch/files/live/sites/irene/files/shared/documents/Publications/Working%20papers/2017/WP17-03_V2.pdf (Last accessed: 20.11.2020).
15. Погода в 243 країнах світу. *rp 5. ua. Розклад погоди.* URL: https://rp5.ua/ (Last accessed: 05.11.2020).

16. A box and whiskers plot (in the style of Tukey). *ggplot2*. URL: https://ggplot2.tidyverse.org/reference/geom_boxplot.html (Last accessed: 01.12.2020).
17. Hyndman R. J., Athanasopoulos G. Forecasting: Principles and Practice. 2nd edition, OTexts : Melbourne, 2018. URL: https://otexts.com/fpp2/ (Last accessed: 05.12.2020).
18. ANOVA in R. *Data Novia*. URL: https://www.datanovia.com/en/lessons/anova-in-r/ (Last accessed: 16.11.2020).
19. Linear Models for Panel Data : package 'plm' / Y. Croissant et al. Published: 13.10.2020. URL: https://cran.r-project.org/web/packages/plm/plm.pdf (Last accessed: 07.12.2020). License: GPL (>= 2).
20. Torres-Reyna O. Getting Started in Fixed/Random Effects Models using R. Princeton University, 2010. URL: https://www.princeton.edu/~otorres/Panel101R.pdf (Last accessed: 06.12.2020).
21. lmtest: Testing Linear Regression Models : A collection of tests, data sets, and examples for diagnostic checking in linear regression models / T. Hothorn et al. Data, published 09.09.2020. URL: https://cran.r-project.org/web/packages/lmtest/index.html (Last accessed: 04.12.2020). License: GPL-2, GPL-3.

ДОДАТКИ

Додаток А

Причинно-наслідкові зв’язки за тестами Грейнджера та Думітреску

Pairwise Granger Causality Tests
Date: 12/20/20 Time: 18:26
Sample: 2017M11 2020M11
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
T does not Granger Cause CONS	245	15.7827	4.E-07
CONS does not Granger Cause T		4.77405	0.0093
HOLIDAY does not Granger Cause CONS	245	9.98944	7.E-05
CONS does not Granger Cause HOLIDAY		5.75436	0.0036
LIGHT does not Granger Cause CONS	245	13.8659	2.E-06
CONS does not Granger Cause LIGHT		4.03218	0.0190
MONTH does not Granger Cause CONS	245	12.3005	8.E-06
CONS does not Granger Cause MONTH		16.2805	2.E-07
WEEKEND does not Granger Cause CONS	245	0.38888	0.6782
CONS does not Granger Cause WEEKEND		0.66428	0.5156
U does not Granger Cause CONS	245	12.3340	8.E-06
CONS does not Granger Cause U		17.3906	9.E-08

Pairwise Granger Causality Tests
Date: 12/20/20 Time: 17:53
Sample: 2017M11 2020M11
Lags: 3

Null Hypothesis:	Obs	F-Statistic	Prob.
T does not Granger Cause CONS	238	12.4707	1.E-07
CONS does not Granger Cause T		1.68756	0.1704
HOLIDAY does not Granger Cause CONS	238	7.87128	5.E-05
CONS does not Granger Cause HOLIDAY		8.95800	1.E-05
LIGHT does not Granger Cause CONS	238	14.1344	2.E-08
CONS does not Granger Cause LIGHT		3.62485	0.0138
MONTH does not Granger Cause CONS	238	11.5084	5.E-07
CONS does not Granger Cause MONTH		11.8388	3.E-07
WEEKEND does not Granger Cause CONS	238	2.43804	0.0653
CONS does not Granger Cause WEEKEND		0.44902	0.7182
U does not Granger Cause CONS	238	8.40699	3.E-05
CONS does not Granger Cause U		8.97607	1.E-05

Продовження Додатку А

Pairwise Dumitrescu Hurlin Panel Causality Tests
Date: 12/20/20 Time: 18:32
Sample: 2017M11 2020M11
Lags: 2

Null Hypothesis:	W-Stat.	Zbar-Stat.	Prob.
T does not homogeneously cause CONS	5.84721	4.25788	2.E-05
CONS does not homogeneously cause T	7.14246	5.74668	9.E-09
HOLIDAY does not homogeneously cause CONS	5.49056	3.84794	0.0001
CONS does not homogeneously cause HOLIDAY	7.38767	6.02853	2.E-09
LIGHT does not homogeneously cause CONS	9.69515	8.68081	0.0000
CONS does not homogeneously cause LIGHT	2.53221	0.44753	0.6545
MONTH does not homogeneously cause CONS	8.11600	6.86570	7.E-12
CONS does not homogeneously cause MONTH	25.9204	27.3305	0.0000
WEEKEND does not homogeneously cause CONS	1.14744	-1.14416	0.2526
CONS does not homogeneously cause WEEKEND	1.58058	-0.64630	0.5181
U does not homogeneously cause CONS	2.80752	0.76399	0.4449
CONS does not homogeneously cause U	13.1509	12.6530	0.0000

Pairwise Dumitrescu Hurlin Panel Causality Tests
Date: 12/20/20 Time: 18:33
Sample: 2017M11 2020M11
Lags: 3

Null Hypothesis:	W-Stat.	Zbar-Stat.	Prob.
T does not homogeneously cause CONS	7.60701	3.95840	8.E-05
CONS does not homogeneously cause T	5.21801	1.79293	0.0730
HOLIDAY does not homogeneously cause CONS	5.65970	2.19329	0.0283
CONS does not homogeneously cause HOLIDAY	11.8005	7.75950	8.E-15
LIGHT does not homogeneously cause CONS	11.4262	7.42021	1.E-13
CONS does not homogeneously cause LIGHT	7.23443	3.62067	0.0003
MONTH does not homogeneously cause CONS	6.97644	3.38683	0.0007
CONS does not homogeneously cause MONTH	37.5816	31.1283	0.0000
WEEKEND does not homogeneously cause CONS	2.54099	-0.63360	0.5263
CONS does not homogeneously cause WEEKEND	1.50673	-1.57109	0.1162
U does not homogeneously cause CONS	2.68952	-0.49897	0.6178
CONS does not homogeneously cause U	12.7760	8.64370	0.0000

Pairwise Dumitrescu Hurlin Panel Causality Tests
Date: 12/20/20 Time: 18:39
Sample: 2017M11 2020M11
Lags: 6

Null Hypothesis:	W-Stat.	Zbar-Stat.	Prob.
T does not homogeneously cause CONS	9.82342	1.66449	0.0960
CONS does not homogeneously cause T	9.65591	1.57377	0.1155
HOLIDAY does not homogeneously cause CONS	10.7749	2.17976	0.0293
CONS does not homogeneously cause HOLIDAY	32.8299	14.1242	0.0000
LIGHT does not homogeneously cause CONS	11.6569	2.65745	0.0079
CONS does not homogeneously cause LIGHT	15.4193	4.69509	3.E-06
MONTH does not homogeneously cause CONS	11.3469	2.48954	0.0128
CONS does not homogeneously cause MONTH	33.4421	14.4558	0.0000
WEEKEND does not homogeneously cause CONS	11.0895	2.35015	0.0188
CONS does not homogeneously cause WEEKEND	7.01278	0.14232	0.8868
U does not homogeneously cause CONS	10.4523	2.00507	0.0450
CONS does not homogeneously cause U	20.8762	7.65042	2.E-14

Pairwise Dumitrescu Hurlin Panel Causality Tests
Date: 12/20/20 Time: 18:40
Sample: 2017M11 2020M11
Lags: 9

Null Hypothesis:	W-Stat.	Zbar-Stat.	Prob.
T does not homogeneously cause CONS	12.3732	0.21739	0.8279
CONS does not homogeneously cause T	17.1328	1.50790	0.1316
HOLIDAY does not homogeneously cause CONS	25.8093	3.86044	0.0001
CONS does not homogeneously cause HOLIDAY	34.3793	6.18412	6.E-10
LIGHT does not homogeneously cause CONS	20.4307	2.40211	0.0163
CONS does not homogeneously cause LIGHT	22.8300	3.05264	0.0023
MONTH does not homogeneously cause CONS	19.8617	2.24782	0.0246
CONS does not homogeneously cause MONTH	41.0865	8.00272	1.E-15
WEEKEND does not homogeneously cause CONS	13.2219	0.44752	0.6545
CONS does not homogeneously cause WEEKEND	16.4319	1.31787	0.1875
U does not homogeneously cause CONS	17.1669	1.51714	0.1292
CONS does not homogeneously cause U	9.87190	-0.46081	0.6449

Додаток Б

Оцінки коефіцієнтів короткострокові динаміки за моделлю ARDL(4, 4, 4, 4) у розрізі обєктів (корпусів)

Корпус Г:

G: Variable	Coefficient	Std. Error	t-Statistic	Prob. *
COINTEQ01	-0.011374	0.000716	-15.89092	0.0005
D(LOGCONS(-1))	-0.241581	0.068735	-3.514676	0.0391
D(LOGCONS(-2))	-0.524610	0.071196	-7.368555	0.0052
D(LOGCONS(-3))	-0.264589	0.069342	-3.815730	0.0317
D(T)	-0.018701	9.11E-05	-205.3013	0.0000
D(T(-1))	-0.014698	8.76E-05	-167.7080	0.0000
D(T(-2))	-0.008438	8.84E-05	-95.41656	0.0000
D(T(-3))	-0.005625	8.02E-05	-70.17787	0.0000
D(WEEKEND)	1.690212	14.01615	0.120590	0.9116
D(WEEKEND(-1))	0.110279	9.965143	0.011067	0.9919
D(WEEKEND(-2))	-2.383999	6.338636	-0.376106	0.7319
D(WEEKEND(-3))	-1.815618	2.433630	-0.746053	0.5098
D(HOLIDAY)	2.632461	20.06067	0.131225	0.9039
D(HOLIDAY(-1))	2.322748	10.75843	0.215900	0.8429
D(HOLIDAY(-2))	-0.224773	6.419911	-0.035012	0.9743
D(HOLIDAY(-3))	-0.999109	3.213569	-0.310903	0.7762
C	0.735025	2.406000	0.305497	0.7800

Корпус НВК:

NVK: Variable	Coefficient	Std. Error	t-Statistic	Prob. *
COINTEQ01	-0.156010	0.038214	-4.082537	0.0265
D(LOGCONS(-1))	-0.614720	0.042372	-14.50773	0.0007
D(LOGCONS(-2))	-0.435007	0.023850	-18.23905	0.0004
D(LOGCONS(-3))	-0.446598	0.026495	-16.85574	0.0005
D(T)	-0.009438	9.39E-05	-100.4876	0.0000
D(T(-1))	-0.013218	8.75E-05	-151.1072	0.0000
D(T(-2))	-0.014983	0.000107	-139.7177	0.0000
D(T(-3))	-0.060461	0.000104	-581.4032	0.0000
D(WEEKEND)	19.51875	22.58998	0.864045	0.4511
D(WEEKEND(-1))	11.98098	15.40977	0.777492	0.4935
D(WEEKEND(-2))	1.987690	7.379294	0.269360	0.8051
D(WEEKEND(-3))	4.591645	2.379158	1.929945	0.1492
D(HOLIDAY)	27.00971	20.38688	1.324857	0.2771
D(HOLIDAY(-1))	20.22934	11.15987	1.812685	0.1675
D(HOLIDAY(-2))	10.79604	4.881997	2.211399	0.1140
D(HOLIDAY(-3))	6.635530	1.739805	3.813950	0.0317
C	10.45781	4.692280	2.228727	0.1121

Корпус С:

C Variable	Coefficient	Std. Error	t-Statistic	Prob. *
COINTEQ01	-0.023721	0.001276	-18.58996	0.0003
D(LOGCONS(-1))	0.139521	0.077104	1.809519	0.1681
D(LOGCONS(-2))	-0.417910	0.036055	-11.59103	0.0014
D(LOGCONS(-3))	-0.067556	0.037044	-1.823666	0.1657
D(T)	-0.006494	4.71E-05	-137.9130	0.0000
D(T(-1))	-0.011132	5.39E-05	-206.7045	0.0000
D(T(-2))	-0.007235	5.60E-05	-129.1816	0.0000
D(T(-3))	-0.001905	8.15E-05	-23.37339	0.0002
D(WEEKEND)	3.323300	9.067458	0.366508	0.7383
D(WEEKEND(-1))	1.062103	5.766660	0.184180	0.8656
D(WEEKEND(-2))	-1.238648	3.449432	-0.359087	0.7433
D(WEEKEND(-3))	-0.178878	1.327373	-0.134761	0.9013
D(HOLIDAY)	4.036068	13.31759	0.303063	0.7816
D(HOLIDAY(-1))	3.223609	6.633348	0.485970	0.6603
D(HOLIDAY(-2))	0.752013	3.567612	0.210789	0.8466
D(HOLIDAY(-3))	-0.042018	1.287775	-0.032628	0.9760
C	1.607682	1.843555	0.872055	0.4474

Корпус Л:

L: Variable	Coefficient	Std. Error	t-Statistic	Prob. *
COINTEQ01	-0.013135	0.000513	-25.61292	0.0001
D(LOGCONS(-1))	-0.162705	0.028223	-5.764893	0.0104
D(LOGCONS(-2))	0.011981	0.030002	0.399331	0.7164
D(LOGCONS(-3))	-0.179058	0.036514	-4.903854	0.0162
D(T)	-0.001923	4.46E-05	-43.07931	0.0000
D(T(-1))	-0.004435	4.32E-05	-102.6649	0.0000
D(T(-2))	0.000631	3.69E-05	17.11053	0.0004
D(T(-3))	-0.004576	4.38E-05	-104.4697	0.0000
D(WEEKEND)	0.683435	5.995857	0.113985	0.9165
D(WEEKEND(-1))	0.651212	5.045055	0.129079	0.9055
D(WEEKEND(-2))	0.046056	3.147995	0.014630	0.9892
D(WEEKEND(-3))	1.291961	1.168898	1.105281	0.3497
D(HOLIDAY)	1.546112	7.383142	0.209411	0.8475
D(HOLIDAY(-1))	1.382972	4.585138	0.301621	0.7826
D(HOLIDAY(-2))	0.879158	2.376869	0.369881	0.7360
D(HOLIDAY(-3))	0.519373	0.943476	0.550489	0.6203
C	0.881856	0.954173	0.924209	0.4235

Корпус БІЦ:

LIB: Variable	Coefficient	Std. Error	t-Statistic	Prob. *
COINTEQ01	-0.031192	0.001747	-17.85152	0.0004
D(LOGCONS(-1))	-0.423020	0.021693	-19.50048	0.0003
D(LOGCONS(-2))	-0.569491	0.017635	-32.29374	0.0001
D(LOGCONS(-3))	-0.555896	0.017669	-31.46200	0.0001
D(T)	-0.035316	4.93E-05	-715.8295	0.0000
D(T(-1))	-0.022799	5.34E-05	-426.8463	0.0000
D(T(-2))	-0.005853	4.97E-05	-117.8813	0.0000
D(T(-3))	-0.030205	5.66E-05	-533.7413	0.0000
D(WEEKEND)	3.307244	9.328847	0.354518	0.7464
D(WEEKEND(-3))	-2.300162	1.279903	-1.797138	0.1702
C	2.057470	1.612874	1.275654	0.2919

Корпус Т:

T: Variable	Coefficient	Std. Error	t-Statistic	Prob. *
COINTEQ01	-0.057121	0.004947	-11.54672	0.0014
D(LOGCONS(-1))	0.282091	0.028621	9.856029	0.0022
D(LOGCONS(-2))	-0.408881	0.020459	-19.98536	0.0003
D(LOGCONS(-3))	-0.224593	0.024192	-9.283687	0.0026
D(T)	-0.013806	5.23E-05	-263.8184	0.0000
D(T(-1))	-0.029698	5.70E-05	-520.9492	0.0000
D(T(-2))	0.014564	5.36E-05	271.6358	0.0000
D(T(-3))	-0.033370	5.29E-05	-630.2602	0.0000
D(WEEKEND(-3))	3.062546	1.621530	1.888677	0.1554
D(HOLIDAY(-3))	1.479329	1.179035	1.254695	0.2984
C	3.734723	1.279903	2.917974	0.0616

Корпус М:

M: Variable	Coefficient	Std. Error	t-Statistic	Prob. *
COINTEQ01	-0.081863	0.009767	-8.381146	0.0036
D(LOGCONS(-1))	-0.189271	0.024023	-7.878600	0.0043
D(LOGCONS(-2))	-0.488025	0.016375	-29.80258	0.0001
D(LOGCONS(-3))	-0.215154	0.018276	-11.77268	0.0013
D(T)	-0.016730	6.02E-05	-278.1155	0.0000
D(T(-1))	-0.019220	5.47E-05	-351.6615	0.0000
D(T(-2))	-0.011896	4.64E-05	-256.2466	0.0000
D(T(-3))	-0.036642	5.97E-05	-614.0085	0.0000
D(WEEKEND)	10.71706	11.31046	0.947536	0.4133
D(WEEKEND(-1))	7.057418	8.051673	0.876516	0.4453
D(WEEKEND(-3))	1.855581	1.480266	1.253546	0.2988
D(HOLIDAY(-3))	1.806509	1.521439	1.187369	0.3205
C	5.369538	1.745253	3.076652	0.0543

Додаток В

Програмна реалізація в RStudio

```
# create a subset without G and C
panel$Object <- as.factor(panel$Object)
levels(panel$Object)
## [1] "C"   "G"   "L"   "Lib" "M"   "NVK" "T"
panel_cg <- subset(panel, Object=='G' | Object=='C')

panel_c <- subset(panel, Object=='C')
panel_g <- subset(panel, Object=='G')

summary(panel_c)
##      Date           Object          Cons              T
##  Length:1106        C  :1106   Min.   : 206.7   Min.   :-15.6875
##  Class :character   G  :   0   1st Qu.: 775.7   1st Qu.:  0.5281
##  Mode  :character   L  :   0   Median :1014.0   Median :  9.3339
##                     Lib:   0   Mean   :1125.1   Mean   :  9.0923
##                     M  :   0   3rd Qu.:1472.8   3rd Qu.: 18.0125
##                     NVK:   0   Max.   :2261.6   Max.   : 26.9375
##                     T  :   0
##        Po              P               Pa                   U
##  Min.   :726.5   Min.   :742.8   Min.   :-2.087500   Min.   : 28.25
##  1st Qu.:742.1   1st Qu.:758.6   1st Qu.:-0.287500   1st Qu.: 61.12
##  Median :745.8   Median :762.1   Median : 0.012500   Median : 73.88
##  Mean   :745.9   Mean   :762.5   Mean   : 0.002563   Mean   : 73.61
##  3rd Qu.:749.7   3rd Qu.:766.4   3rd Qu.: 0.287500   3rd Qu.: 87.72
##  Max.   :764.2   Max.   :782.2   Max.   : 1.687500   Max.   :100.00
##
##        Ff        Weekend Holiday      month            light
##  Min.   :0.375   0:796   0:1063   Min.   : 1.000   Min.   :0.2917
##  1st Qu.:2.375   1:310   1:  43   1st Qu.: 4.000   1st Qu.:0.3750
##  Median :3.250                    Median : 7.000   Median :0.5037
##  Mean   :3.467                    Mean   : 6.562   Mean   :0.4838
##  3rd Qu.:4.500                    3rd Qu.:10.000   3rd Qu.:0.6095
##  Max.   :9.500                    Max.   :12.000   Max.   :0.6805
##
sd(panel_c$Cons)
## [1] 435.0545
summary(panel_g)
##      Date           Object          Cons              T
##  Length:1106        C  :   0   Min.   : 257.7   Min.   :-15.6875
##  Class :character   G  :1106   1st Qu.: 504.3   1st Qu.:  0.5281
##  Mode  :character   L  :   0   Median : 834.1   Median :  9.3339
##                     Lib:   0   Mean   : 967.0   Mean   :  9.0923
##                     M  :   0   3rd Qu.:1412.6   3rd Qu.: 18.0125
##                     NVK:   0   Max.   :2279.9   Max.   : 26.9375
##                     T  :   0
##        Po              P               Pa                   U
##  Min.   :726.5   Min.   :742.8   Min.   :-2.087500   Min.   : 28.25
##  1st Qu.:742.1   1st Qu.:758.6   1st Qu.:-0.287500   1st Qu.: 61.12
##  Median :745.8   Median :762.1   Median : 0.012500   Median : 73.88
##  Mean   :745.9   Mean   :762.5   Mean   : 0.002563   Mean   : 73.61
##  3rd Qu.:749.7   3rd Qu.:766.4   3rd Qu.: 0.287500   3rd Qu.: 87.72
##  Max.   :764.2   Max.   :782.2   Max.   : 1.687500   Max.   :100.00
##
##        Ff        Weekend Holiday      month            light
##  Min.   :0.375   0:796   0:1063   Min.   : 1.000   Min.   :0.2917
##  1st Qu.:2.375   1:310   1:  43   1st Qu.: 4.000   1st Qu.:0.3750
##  Median :3.250                    Median : 7.000   Median :0.5037
##  Mean   :3.467                    Mean   : 6.562   Mean   :0.4838
##  3rd Qu.:4.500                    3rd Qu.:10.000   3rd Qu.:0.6095
##  Max.   :9.500                    Max.   :12.000   Max.   :0.6805
##
sd(panel_g$Cons)
```

```
## [1] 525.1766
# create a subset without G and C
attach(panel)
## The following object is masked from package:base:
##
##     T
panel_s1 <- panel[which(Object!='G'),]
detach(panel)
attach(panel_s1)
## The following object is masked from package:base:
##
##     T
panel_s1 <- panel_s1[which(Object!='C'),]
detach(panel_s1)

# remove NA rows
panel_s1 <- na.omit(panel_s1)

# check whether the data are balanced:
suppressMessages(library(plm)) # to remove note about dependencies
is.pbalanced(panel_s1) # are the data balanced?
## [1] TRUE
summary(panel_s1)
##       Date           Object          Cons               T
##  Length:5530        C  :   0   Min.   :   9.04   Min.   :-15.688
##  Class :character   G  :   0   1st Qu.:  69.22   1st Qu.:  0.525
##  Mode  :character   L  :1106   Median : 152.37   Median :  9.334
##                     Lib:1106   Mean   : 238.94   Mean   :  9.092
##                     M  :1106   3rd Qu.: 379.77   3rd Qu.: 18.012
##                     NVK:1106   Max.   :1135.50   Max.   : 26.938
##                     T  :1106
##        Po              P               Pa                  U
##  Min.   :726.5   Min.   :742.8   Min.   :-2.087500   Min.   : 28.25
##  1st Qu.:742.1   1st Qu.:758.6   1st Qu.:-0.287500   1st Qu.: 61.12
##  Median :745.8   Median :762.1   Median : 0.012500   Median : 73.88
##  Mean   :745.9   Mean   :762.5   Mean   : 0.002563   Mean   : 73.61
##  3rd Qu.:749.7   3rd Qu.:766.4   3rd Qu.: 0.287500   3rd Qu.: 87.75
##  Max.   :764.2   Max.   :782.2   Max.   : 1.687500   Max.   :100.00
##
##        Ff        Weekend  Holiday      month            light
##  Min.   :0.375   0:3980   0:5315   Min.   : 1.000   Min.   :0.2917
##  1st Qu.:2.375   1:1550   1: 215   1st Qu.: 4.000   1st Qu.:0.3750
##  Median :3.250                     Median : 7.000   Median :0.5037
##  Mean   :3.467                     Mean   : 6.562   Mean   :0.4838
##  3rd Qu.:4.500                     3rd Qu.:10.000   3rd Qu.:0.6095
##  Max.   :9.500                     Max.   :12.000   Max.   :0.6805
##
sd(panel_s1$Cons)
## [1] 219.3772
library(ggplot2)
# Check Consumption distribution all buildings
ggplot(panel, aes(Cons)) +  geom_histogram()
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
# Check Consumption distribution buildings C and G
ggplot(panel_cg, aes(Cons)) +  geom_histogram()
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
qplot(panel_c$Cons)
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
qplot(panel_g$Cons)
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
# Check Consumption distribution other buildings without C and G
ggplot(panel_s1, aes(Cons)) +  geom_histogram()
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
# Check Consumption distribution by Weekend and by Holiday

ggplot(panel, aes(Cons, fill=Weekend)) +
  geom_histogram()
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
ggplot(panel, aes(Cons, fill=Holiday)) +
  geom_histogram()
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
# Create boxplot by Object

panel$Object <- as.factor(panel$Object)

ggplot(panel, aes(Cons, fill=Object)) +
  geom_boxplot()
```

```
# We can see that the consumption differs greatly by object, having higher values of
bigger buildings - G, C

# Create boxplot by Month

panel$month <- as.factor(panel$month)

ggplot(panel, aes(Cons, fill=month)) +
  geom_boxplot()
```

```
# log transform variable Cons and add to df
library(dplyr)
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:plm':
##
##     between, lag, lead
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
panel = panel %>%
dplyr::mutate(logcons = log(panel$Cons))

library(dplyr)
panel_s1 = panel_s1 %>%
dplyr::mutate(logcons = log(panel_s1$Cons))

library(dplyr)
panel_cg = panel_cg %>%
dplyr::mutate(logcons = log(panel_cg$Cons))
# The Dickey-Fuller test to check for stochastic trends. The null hypothesis is that
the series has a unit root (i.e. non-stationary). If unit root is present you can take
the first difference of the variable.

library(plm)
```

```
library(tseries)
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
library("punitroots")
## Loading required package: CADFtest
## Warning: package 'CADFtest' was built under R version 4.0.2
## Loading required package: dynlm
## Warning: package 'dynlm' was built under R version 4.0.2
## Loading required package: zoo
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
## Loading required package: sandwich
## Loading required package: urca
## Warning: package 'urca' was built under R version 4.0.2
## Registered S3 methods overwritten by 'CADFtest':
##   method     from
##   bread.mlm  sandwich
##   estfun.mlm sandwich
## Loading required package: fUnitRoots
## Warning: package 'fUnitRoots' was built under R version 4.0.2
## Loading required package: timeDate
## Loading required package: timeSeries
## No methods found in package 'timeDate' for request: '[<-' when loading 'timeSeries'
##
## Attaching package: 'timeSeries'
## The following object is masked from 'package:zoo':
##
##     time<-
## Loading required package: fBasics
## Warning: package 'fBasics' was built under R version 4.0.2
##
## Attaching package: 'fUnitRoots'
## The following objects are masked from 'package:urca':
##
##     punitroot, qunitroot, unitrootTable
# Create a panel data frame
Panel.set <- pdata.frame(panel, index = c('Object', 'Date'))

# Augmented Dickey-Fuller Test for Cons variable
adf.test(panel$logcons)
## Warning in adf.test(panel$logcons): p-value smaller than printed p-value
##
##  Augmented Dickey-Fuller Test
##
## data:  panel$logcons
## Dickey-Fuller = -4.0982, Lag order = 19, p-value = 0.01
## alternative hypothesis: stationary
adf.test(panel_c$Cons)
##
##  Augmented Dickey-Fuller Test
##
## data:  panel_c$Cons
## Dickey-Fuller = -2.4891, Lag order = 10, p-value = 0.3713
## alternative hypothesis: stationary
adf.test(panel_g$Cons)
## Warning in adf.test(panel_g$Cons): p-value smaller than printed p-value
##
##  Augmented Dickey-Fuller Test
##
## data:  panel_g$Cons
## Dickey-Fuller = -4.1252, Lag order = 10, p-value = 0.01
## alternative hypothesis: stationary
# Unit Root Tests For Panel Data
purtest(logcons~trend,data=panel,index=c('Object', 'Date'),pmax=3,test='levinlin')
## Warning in selectT(l, theTs): the time series is long
```

```
##
##  Levin-Lin-Chu Unit-Root Test (ex. var.: Individual Intercepts and
##  Trend)
##
## data:  logcons ~ trend
## z = -41.449, p-value < 2.2e-16
## alternative hypothesis: stationarity
purtest(logcons~trend,data=panel,index=c('Object', 'Date'),pmax=3,test='hadri')
##
##  Hadri Test (ex. var.: Individual Intercepts and Trend) (Heterosked.
##  Consistent)
##
## data:  logcons ~ trend
## z = 101.4, p-value < 2.2e-16
## alternative hypothesis: at least one series has a unit root
adf.test(panel_s1$logcons)
## Warning in adf.test(panel_s1$logcons): p-value smaller than printed p-value
##
##  Augmented Dickey-Fuller Test
##
## data:  panel_s1$logcons
## Dickey-Fuller = -4.5191, Lag order = 17, p-value = 0.01
## alternative hypothesis: stationary
purtest(logcons~trend,data=panel_s1,index=c('Object', 'Date'),pmax=3,test='levinlin')
## Warning in selectT(l, theTs): the time series is long
##
##  Levin-Lin-Chu Unit-Root Test (ex. var.: Individual Intercepts and
##  Trend)
##
## data:  logcons ~ trend
## z = -35.327, p-value < 2.2e-16
## alternative hypothesis: stationarity
purtest(logcons~trend,data=panel_s1,index=c('Object', 'Date'),pmax=3,test='hadri')
##
##  Hadri Test (ex. var.: Individual Intercepts and Trend) (Heterosked.
##  Consistent)
##
## data:  logcons ~ trend
## z = 80.109, p-value < 2.2e-16
## alternative hypothesis: at least one series has a unit root
adf.test(panel_cg$logcons)
##
##  Augmented Dickey-Fuller Test
##
## data:  panel_cg$logcons
## Dickey-Fuller = -3.1628, Lag order = 13, p-value = 0.09432
## alternative hypothesis: stationary
purtest(logcons~trend,data=panel_cg,index=c('Object', 'Date'),pmax=3,test='levinlin')
## Warning in selectT(l, theTs): the time series is long
##
##  Levin-Lin-Chu Unit-Root Test (ex. var.: Individual Intercepts and
##  Trend)
##
## data:  logcons ~ trend
## z = -21.686, p-value < 2.2e-16
## alternative hypothesis: stationarity
purtest(logcons~trend,data=panel_cg,index=c('Object', 'Date'),pmax=3,test='hadri')
##
##  Hadri Test (ex. var.: Individual Intercepts and Trend) (Heterosked.
##  Consistent)
##
## data:  logcons ~ trend
## z = 63.032, p-value < 2.2e-16
## alternative hypothesis: at least one series has a unit root
# If p-value < 0.05 then no unit roots present
# Consumption series are stationary by the ADF and Levin-Lin-Chu tests, but
unstationary by Hadri test
library(car)
## Loading required package: carData
```

```
##
## Attaching package: 'car'
## The following object is masked from 'package:fBasics':
##
##     densityPlot
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
# Explore panel data

par(mar=c(1,1,1,1))

# Heterogeneity across entities
library(gplots)
```

```
##
## Attaching package: 'gplots'
## The following object is masked from 'package:stats':
##
##     lowess
```

```
plotmeans(Cons ~ Object, main="", data=panel,las=2, xlab = "", ylab = "Mean
electricity consumption")
```

```
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```

```
plotmeans(Cons ~ Date, main="", data=panel,las=2, xlab = "", ylab = "Mean electricity
consumption")
```

```
plotmeans(Cons ~ month, main="", data=panel,las=2, xlab = "", ylab = "Mean electricity
consumption")
```

```
plotmeans(Cons ~ Object, main="", data=panel_s1,las=2, xlab = "", ylab = "")
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```

```
plotmeans(Cons ~ Date, main="", data=panel_s1,las=2, xlab = "", ylab = "")
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped

## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, : zero-
## length arrow is of indeterminate angle and so skipped
```

```
# Scatterpolts
library(GGally)
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
panel$month <- as.numeric(panel$month)
ggpairs(panel,columns = c(3,4,5,7,8,9,12,13))
```

```
# H0: mean electricity consumption(Weekend =0) = mean consumption(Weekend =1)

aggregate(x = panel$logcons,
          by = list(panel$Weekend),
          FUN = mean)
##   Group.1        x
## 1       0 5.781454
## 2       1 4.903459
aov<- aov(logcons ~ Weekend, data=panel)
summary(aov)
##               Df Sum Sq Mean Sq F value Pr(>F)
## Weekend        1   1204  1203.9   904.6 <2e-16 ***
## Residuals   7740  10301     1.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
plot(aov)
```

```
# H0: mean electricity consumption(Holiday =0) = mean consumption(Holiday=1)
```

```
aggregate(x = panel$logcons,
          by = list(panel$Holiday),
          FUN = mean)
##   Group.1        x
## 1       0 5.562312
## 2       1 4.869102
aov1<- aov(logcons ~ Holiday, data=panel)
summary(aov1)
##               Df Sum Sq Mean Sq F value Pr(>F)
## Holiday        1    139  139.02   94.67 <2e-16 ***
## Residuals   7740  11366    1.47
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
plot(aov1)
```

```
# H0: mean electricity consumption(Object 1) = mean electricity consumption(Object 2)
= mean electricity consumption(Object 3)=...
group_by(panel, Object) %>%
  summarise(mean = mean(logcons),
    sd = sd(logcons))
## `summarise()` ungrouping output (override with `.groups` argument)
## # A tibble: 7 x 3
##   Object  mean    sd
##   <fct>  <dbl> <dbl>
## 1 C       6.95 0.394
## 2 G       6.71 0.592
## 3 L       5.94 0.836
## 4 Lib     5.27 0.766
## 5 M       4.41 0.685
## 6 NVK     5.54 0.501
## 7 T       3.93 0.650
aov2<- aov(logcons ~ Object, data=panel)
summary(aov2)
##               Df Sum Sq Mean Sq F value Pr(>F)
## Object         6   8260  1376.7    3283 <2e-16 ***
## Residuals   7735   3244     0.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
plot(aov2)
```

```
# Post Hoc Tests Tukey Tukey's Honest Significant Difference (HSD) multiple pairwise-
comparisons to determine if the mean difference between specific pairs of group are
statistically significant

TukeyHSD(aov2)
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = logcons ~ Object, data = panel)
##
## $Object
##               diff        lwr        upr p adj
## G-C     -0.2385469 -0.3197641 -0.1573297     0
## L-C     -1.0107495 -1.0919667 -0.9295323     0
## Lib-C   -1.6837879 -1.7650051 -1.6025707     0
## M-C     -2.5376071 -2.6188243 -2.4563899     0
## NVK-C   -1.4074486 -1.4886658 -1.3262314     0
## T-C     -3.0232510 -3.1044682 -2.9420338     0
## L-G     -0.7722026 -0.8534198 -0.6909854     0
## Lib-G   -1.4452410 -1.5264582 -1.3640238     0
## M-G     -2.2990602 -2.3802774 -2.2178430     0
## NVK-G   -1.1689017 -1.2501189 -1.0876845     0
## T-G     -2.7847041 -2.8659213 -2.7034869     0
## Lib-L   -0.6730384 -0.7542556 -0.5918212     0
## M-L     -1.5268576 -1.6080748 -1.4456404     0
## NVK-L   -0.3966991 -0.4779163 -0.3154819     0
```

```
## T-L     -2.0125015 -2.0937187 -1.9312843     0
## M-Lib   -0.8538192 -0.9350364 -0.7726020     0
## NVK-Lib  0.2763393  0.1951221  0.3575565     0
## T-Lib   -1.3394631 -1.4206803 -1.2582459     0
## NVK-M    1.1301585  1.0489413  1.2113757     0
## T-M     -0.4856439 -0.5668611 -0.4044267     0
## T-NVK   -1.6158024 -1.6970196 -1.5345853     0
#  Levene?s test to check the homogeneity of variances
library(car)
leveneTest(logcons ~ Object, data=panel)
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value    Pr(>F)
## group    6  89.253 < 2.2e-16 ***
##       7735
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
# p-value < 0.05 we can't assume the homogeneity of variances in the different
treatment groups

# Kruskal-Wallis rank sum test, which can be used when ANNOVA assumptions are not met
kruskal.test(logcons ~ Object, data=panel)
##
##  Kruskal-Wallis rank sum test
##
## data:  logcons by Object
## Kruskal-Wallis chi-squared = 5676.8, df = 6, p-value < 2.2e-16
# H0: mean electricity consumption(Month 1) = mean electricity consumption(Month 2) =
mean electricity consumption(Month 3)=...
group_by(panel, month) %>%
  summarise(mean = mean(logcons),
    sd = sd(logcons))
## `summarise()` ungrouping output (override with `.groups` argument)
## # A tibble: 12 x 3
##    month  mean    sd
##    <dbl> <dbl> <dbl>
##  1     1  5.88  1.12
##  2     2  5.90  1.11
##  3     3  5.66  1.14
##  4     4  5.45  1.17
##  5     5  5.29  1.17
##  6     6  5.20  1.27
##  7     7  5.06  1.29
##  8     8  5.04  1.30
##  9     9  5.40  1.21
## 10    10  5.68  1.14
## 11    11  5.95  1.07
## 12    12  5.91  1.09
aov3<- aov(logcons ~ month, data=panel)
summary(aov3)
##               Df Sum Sq Mean Sq F value Pr(>F)
## month          1      0  0.1603   0.108  0.743
## Residuals   7740  11504  1.4864
library(lmtest)

# Temperature Granger causes Cons: past values of temperature improve prediction of
Cons (compared to only past values of Cons) at lag=1

grangertest(logcons ~ T, order = 1, data = panel)
## Granger causality test
##
## Model 1: logcons ~ Lags(logcons, 1:1) + Lags(T, 1:1)
## Model 2: logcons ~ Lags(logcons, 1:1)
##   Res.Df Df      F    Pr(>F)
## 1   7738
## 2   7739 -1 31.352 2.226e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
grangertest(logcons ~ T, order = 2, data = panel)
## Granger causality test
```

```
##
## Model 1: logcons ~ Lags(logcons, 1:2) + Lags(T, 1:2)
## Model 2: logcons ~ Lags(logcons, 1:2)
##   Res.Df Df      F    Pr(>F)
## 1   7735
## 2   7737 -2 15.704 1.561e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
grangertest(logcons ~ T, order = 3, data = panel)
## Granger causality test
##
## Model 1: logcons ~ Lags(logcons, 1:3) + Lags(T, 1:3)
## Model 2: logcons ~ Lags(logcons, 1:3)
##   Res.Df Df      F   Pr(>F)
## 1   7732
## 2   7735 -3 5.6344 0.000746 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
grangertest(logcons ~ T, order = 4, data = panel)
## Granger causality test
##
## Model 1: logcons ~ Lags(logcons, 1:4) + Lags(T, 1:4)
## Model 2: logcons ~ Lags(logcons, 1:4)
##   Res.Df Df      F   Pr(>F)
## 1   7729
## 2   7733 -4 3.9636 0.003241 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
grangertest(logcons ~ T, order = 5, data = panel)
## Granger causality test
##
## Model 1: logcons ~ Lags(logcons, 1:5) + Lags(T, 1:5)
## Model 2: logcons ~ Lags(logcons, 1:5)
##   Res.Df Df      F  Pr(>F)
## 1   7726
## 2   7731 -5 2.1281 0.05911 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
grangertest(logcons ~ T, order = 6, data = panel)
## Granger causality test
##
## Model 1: logcons ~ Lags(logcons, 1:6) + Lags(T, 1:6)
## Model 2: logcons ~ Lags(logcons, 1:6)
##   Res.Df Df      F   Pr(>F)
## 1   7723
## 2   7729 -6 2.9977 0.006308 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
grangertest(logcons ~ T, order = 7, data = panel)
## Granger causality test
##
## Model 1: logcons ~ Lags(logcons, 1:7) + Lags(T, 1:7)
## Model 2: logcons ~ Lags(logcons, 1:7)
##   Res.Df Df      F    Pr(>F)
## 1   7720
## 2   7727 -7 6.3367 1.916e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
# Temperature Granger causes Cons: past values of temperature improve prediction of
Cons (compared to only past values of Cons) at lag=30

grangertest(logcons ~ light, order = 30, data = panel)
## Granger causality test
##
## Model 1: logcons ~ Lags(logcons, 1:30) + Lags(light, 1:30)
## Model 2: logcons ~ Lags(logcons, 1:30)
##   Res.Df  Df      F    Pr(>F)
## 1   7651
## 2   7681 -30 6.7386 < 2.2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
grangertest(T ~ logcons, order = 30, data = panel)
## Granger causality test
##
## Model 1: T ~ Lags(T, 1:30) + Lags(logcons, 1:30)
## Model 2: T ~ Lags(T, 1:30)
##   Res.Df  Df      F   Pr(>F)
## 1   7651
## 2   7681 -30 3.1094 2.38e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
grangertest(logcons ~ light, order = 1, data = panel)
## Granger causality test
##
## Model 1: logcons ~ Lags(logcons, 1:1) + Lags(light, 1:1)
## Model 2: logcons ~ Lags(logcons, 1:1)
##   Res.Df Df      F    Pr(>F)
## 1   7738
## 2   7739 -1 33.829 6.256e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
grangertest(logcons ~ light, order = 2, data = panel)
## Granger causality test
##
## Model 1: logcons ~ Lags(logcons, 1:2) + Lags(light, 1:2)
## Model 2: logcons ~ Lags(logcons, 1:2)
##   Res.Df Df      F    Pr(>F)
## 1   7735
## 2   7737 -2 16.994 4.322e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
grangertest(logcons ~ light, order = 3, data = panel)
## Granger causality test
##
## Model 1: logcons ~ Lags(logcons, 1:3) + Lags(light, 1:3)
## Model 2: logcons ~ Lags(logcons, 1:3)
##   Res.Df Df      F    Pr(>F)
## 1   7732
## 2   7735 -3 6.1182 0.0003753 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
grangertest(logcons ~ light, order = 4, data = panel)
## Granger causality test
##
## Model 1: logcons ~ Lags(logcons, 1:4) + Lags(light, 1:4)
## Model 2: logcons ~ Lags(logcons, 1:4)
##   Res.Df Df      F   Pr(>F)
## 1   7729
## 2   7733 -4 4.1207 0.002452 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
grangertest(logcons ~ light, order = 5, data = panel)
## Granger causality test
##
## Model 1: logcons ~ Lags(logcons, 1:5) + Lags(light, 1:5)
## Model 2: logcons ~ Lags(logcons, 1:5)
##   Res.Df Df     F  Pr(>F)
## 1   7726
## 2   7731 -5 2.234 0.04822 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
grangertest(logcons ~ light, order = 6, data = panel)
## Granger causality test
##
## Model 1: logcons ~ Lags(logcons, 1:6) + Lags(light, 1:6)
## Model 2: logcons ~ Lags(logcons, 1:6)
##   Res.Df Df      F Pr(>F)
## 1   7723
```

```
## 2   7729 -6 1.1542  0.328
grangertest(logcons ~ light, order = 7, data = panel)
## Granger causality test
##
## Model 1: logcons ~ Lags(logcons, 1:7) + Lags(light, 1:7)
## Model 2: logcons ~ Lags(logcons, 1:7)
##   Res.Df Df      F  Pr(>F)
## 1   7720
## 2   7727 -7 2.4838 0.01515 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
grangertest(logcons ~ light, order = 30, data = panel)
## Granger causality test
##
## Model 1: logcons ~ Lags(logcons, 1:30) + Lags(light, 1:30)
## Model 2: logcons ~ Lags(logcons, 1:30)
##   Res.Df  Df      F    Pr(>F)
## 1   7651
## 2   7681 -30 6.7386 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
grangertest(light ~ logcons, order = 30, data = panel)
## Granger causality test
##
## Model 1: light ~ Lags(light, 1:30) + Lags(logcons, 1:30)
## Model 2: light ~ Lags(light, 1:30)
##   Res.Df  Df      F    Pr(>F)
## 1   7651
## 2   7681 -30 8.2616 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
grangertest(logcons ~ U, order = 1, data = panel)
## Granger causality test
##
## Model 1: logcons ~ Lags(logcons, 1:1) + Lags(U, 1:1)
## Model 2: logcons ~ Lags(logcons, 1:1)
##   Res.Df Df    F   Pr(>F)
## 1   7738
## 2   7739 -1 14.7 0.000127 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
grangertest(logcons ~ U, order = 2, data = panel)
## Granger causality test
##
## Model 1: logcons ~ Lags(logcons, 1:2) + Lags(U, 1:2)
## Model 2: logcons ~ Lags(logcons, 1:2)
##   Res.Df Df      F   Pr(>F)
## 1   7735
## 2   7737 -2 16.759 5.46e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
grangertest(logcons ~ U, order = 3, data = panel)
## Granger causality test
##
## Model 1: logcons ~ Lags(logcons, 1:3) + Lags(U, 1:3)
## Model 2: logcons ~ Lags(logcons, 1:3)
##   Res.Df Df      F    Pr(>F)
## 1   7732
## 2   7735 -3 7.3196 6.748e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
grangertest(logcons ~ U, order = 4, data = panel)
## Granger causality test
##
## Model 1: logcons ~ Lags(logcons, 1:4) + Lags(U, 1:4)
## Model 2: logcons ~ Lags(logcons, 1:4)
##   Res.Df Df      F    Pr(>F)
## 1   7729
## 2   7733 -4 5.2578 0.0003159 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
grangertest(logcons ~ U, order = 5, data = panel)
## Granger causality test
##
## Model 1: logcons ~ Lags(logcons, 1:5) + Lags(U, 1:5)
## Model 2: logcons ~ Lags(logcons, 1:5)
##   Res.Df Df     F   Pr(>F)
## 1   7726
## 2   7731 -5 9.117 1.17e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
grangertest(logcons ~ U, order = 6, data = panel)
## Granger causality test
##
## Model 1: logcons ~ Lags(logcons, 1:6) + Lags(U, 1:6)
## Model 2: logcons ~ Lags(logcons, 1:6)
##   Res.Df Df     F    Pr(>F)
## 1   7723
## 2   7729 -6 5.011 3.904e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
grangertest(logcons ~ U, order = 7, data = panel)
## Granger causality test
##
## Model 1: logcons ~ Lags(logcons, 1:7) + Lags(U, 1:7)
## Model 2: logcons ~ Lags(logcons, 1:7)
##   Res.Df Df      F    Pr(>F)
## 1   7720
## 2   7727 -7 6.8217 4.224e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
grangertest(logcons ~ U, order = 30, data = panel)
## Granger causality test
##
## Model 1: logcons ~ Lags(logcons, 1:30) + Lags(U, 1:30)
## Model 2: logcons ~ Lags(logcons, 1:30)
##   Res.Df  Df      F    Pr(>F)
## 1   7651
## 2   7681 -30 5.4008 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
grangertest(U ~ logcons, order = 30, data = panel)
## Granger causality test
##
## Model 1: U ~ Lags(U, 1:30) + Lags(logcons, 1:30)
## Model 2: U ~ Lags(U, 1:30)
##   Res.Df  Df     F    Pr(>F)
## 1   7651
## 2   7681 -30 5.364 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Specific Granger test for panel data - a combination of Granger tests performed per
individual.
# H0: 'x' does not Granger cause 'y' for all objects
# The formula describes the direction of the (panel) Granger causation where y ~ x
means "x (panel) Granger causes y".
# pgrangertest(formula, data, test = c("Ztilde", "Zbar", "Wbar"), order = 1L, index =
NULL)

#By setting argument test to either "Ztilde" (default) or "Zbar", two different
statistics can be requested. "Ztilde" gives the standardised statistic recommended by
Dumitrescu/Hurlin (2012) for fixed T samples. If set to "Wbar", the intermediate Wbar
statistic (average of individual Granger chi-square statistics) is given which is used
to derive the other two.

#The Zbar statistic is not suitable for unbalanced panels. For the Wbar statistic, no
p-value is available.
```

```
#The implementation uses lmtest::grangertest() from package lmtest to perform the
individual Granger tests.

#data("panel", package = "plm")
#pgrangertest(logcons ~ T, data = panel)

# varying lag order (last individual lag order 3, others lag order 2)
#pgrangertest(logcons ~ T, data = panel)
# ! Regular OLS regression does not consider heterogeneity across groups or time

# Base OLS
ols <-lm(logcons ~ T + Weekend + Holiday, data=panel)
summary(ols)
##
## Call:
## lm(formula = logcons ~ T + Weekend + Holiday, data = panel)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.76428 -0.89829  0.03659  1.01621  2.34405
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.10607    0.01901  321.17   <2e-16 ***
## T           -0.02987    0.00126  -23.71   <2e-16 ***
## Weekend1    -0.93243    0.02805  -33.25   <2e-16 ***
## Holiday1    -0.97137    0.06516  -14.91   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.1 on 7738 degrees of freedom
## Multiple R-squared:  0.1868, Adjusted R-squared:  0.1865
## F-statistic: 592.4 on 3 and 7738 DF,  p-value: < 2.2e-16
# OLS + add. variable month
ols0 <-lm(logcons ~ T + Weekend + Holiday + month, data=panel)
summary(ols0)
##
## Call:
## lm(formula = logcons ~ T + Weekend + Holiday + month, data = panel)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.78273 -0.89479  0.02997  1.01306  2.35746
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.995947   0.028959 207.048  < 2e-16 ***
## T           -0.031191   0.001285 -24.278  < 2e-16 ***
## Weekend1    -0.933140   0.028002 -33.324  < 2e-16 ***
## Holiday1    -0.958744   0.065110 -14.725  < 2e-16 ***
## month        0.018565   0.003687   5.035 4.89e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.098 on 7737 degrees of freedom
## Multiple R-squared:  0.1894, Adjusted R-squared:  0.189
## F-statistic:   452 on 4 and 7737 DF,  p-value: < 2.2e-16
par(mfrow=c(2,2))
plot(ols0)
```

```
# OLS + add. variable month
ols1 <-lm(logcons ~ T + Weekend + Holiday + month + U, data=panel)
summary(ols1)
##
## Call:
## lm(formula = logcons ~ T + Weekend + Holiday + month + U, data = panel)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.79554 -0.89977  0.03594  1.02385  2.33339
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.6753589  0.0792485  71.615  < 2e-16 ***
## T           -0.0272213  0.0015752 -17.281  < 2e-16 ***
## Weekend1    -0.9392743  0.0280056 -33.539  < 2e-16 ***
## Holiday1    -0.9594622  0.0650352 -14.753  < 2e-16 ***
## month        0.0147424  0.0037866   3.893 9.97e-05 ***
## U            0.0042295  0.0009734   4.345 1.41e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.097 on 7736 degrees of freedom
## Multiple R-squared:  0.1914, Adjusted R-squared:  0.1909
## F-statistic: 366.2 on 5 and 7736 DF,  p-value: < 2.2e-16
# OLS + light instead of T
ols2 <-lm(logcons ~ light + Weekend + Holiday + month, data=panel)
summary(ols2)
##
## Call:
## lm(formula = logcons ~ light + Weekend + Holiday + month, data = panel)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.85302 -0.89971  0.04135  1.02047  2.38778
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.152304   0.061388 116.509  < 2e-16 ***
## light       -2.530646   0.104523 -24.211  < 2e-16 ***
## Weekend1    -0.932320   0.028008 -33.288  < 2e-16 ***
## Holiday1    -0.987253   0.065129 -15.158  < 2e-16 ***
## month       -0.014164   0.003659  -3.871 0.000109 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.098 on 7737 degrees of freedom
## Multiple R-squared:  0.1891, Adjusted R-squared:  0.1887
## F-statistic: 451.1 on 4 and 7737 DF,  p-value: < 2.2e-16
par(mfrow=c(2,2))
plot(ols2)
```

```
# anova test to check whether ols or ols1 is better
anova(ols, ols0, ols1)
## Analysis of Variance Table
##
## Model 1: logcons ~ T + Weekend + Holiday
## Model 2: logcons ~ T + Weekend + Holiday + month
## Model 3: logcons ~ T + Weekend + Holiday + month + U
##   Res.Df    RSS Df Sum of Sq     F   Pr(>F)
## 1   7738 9355.9
## 2   7737 9325.3  1    30.556 25.41 4.74e-07 ***
## 3   7736 9302.6  1    22.704 18.88 1.41e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
# ols1 is better

library(apsrtable)
apsrtable(ols, ols0, ols1, ols2, model.names = c("OLS", "OLS0", "OLS1", "OLS2"))
## \begin{table}[!ht]
## \caption{}
## \label{}
## \begin{tabular}{ l D{.}{.}{2}D{.}{.}{2}D{.}{.}{2}D{.}{.}{2} }
## \hline
```

```
##   & \multicolumn{ 1 }{ c }{ OLS } & \multicolumn{ 1 }{ c }{ OLS0 } & \multicolumn{
1 }{ c }{ OLS1 } & \multicolumn{ 1 }{ c }{ OLS2 } \\ \hline
##  %            & OLS       & OLS0      & OLS1      & OLS2     \\
## (Intercept) & 6.11 ^*  & 6.00 ^*  & 5.68 ^*  & 7.15 ^* \\
##             & (0.02)    & (0.03)    & (0.08)    & (0.06)   \\
## T           & -0.03 ^* & -0.03 ^* & -0.03 ^* &          \\
##             & (0.00)    & (0.00)    & (0.00)    &          \\
## Weekend1    & -0.93 ^* & -0.93 ^* & -0.94 ^* & -0.93 ^*\\
##             & (0.03)    & (0.03)    & (0.03)    & (0.03)   \\
## Holiday1    & -0.97 ^* & -0.96 ^* & -0.96 ^* & -0.99 ^*\\
##             & (0.07)    & (0.07)    & (0.07)    & (0.07)   \\
## month       &           & 0.02 ^*  & 0.01 ^*  & -0.01 ^*\\
##             &           & (0.00)    & (0.00)    & (0.00)   \\
## U           &           &           & 0.00 ^*  &          \\
##             &           &           & (0.00)    &          \\
## light       &           &           &           & -2.53 ^*\\
##             &           &           &           & (0.10)    \\
##  $N$         & 7742      & 7742      & 7742      & 7742     \\
## $R^2$       & 0.19      & 0.19      & 0.19      & 0.19     \\
## adj. $R^2$  & 0.19      & 0.19      & 0.19      & 0.19     \\
## Resid. sd   & 1.10      & 1.10      & 1.10      & 1.10      \\ \hline
##  \multicolumn{5}{l}{\footnotesize{Standard errors in parentheses}}\\
## \multicolumn{5}{l}{\footnotesize{$^*$ indicates significance at $p< 0.05 $}}
## \end{tabular}
##  \end{table}
```

```
# make prediction with OLS

plot(panel$T, panel$logcons, pch=19, xlab="temperature", ylab="consumption")
abline(lm(logcons ~ T + Weekend + Holiday + month, data=panel),lwd=3, col="red")
## Warning in abline(lm(logcons ~ T + Weekend + Holiday + month, data = panel), :
## only using the first two of 5 regression coefficients
```

```
#Fixed effects using Least squares dummy variable model (with factor(obj))
# pooled data
fixed.dum <-lm(logcons ~ T + Weekend + Holiday + month + factor(Object), data=panel)
summary(fixed.dum)
##
## Call:
## lm(formula = logcons ~ T + Weekend + Holiday + month + factor(Object),
##     data = panel)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.77538 -0.22265  0.03112  0.24888  1.47253
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         7.4104311  0.0142339  520.62   <2e-16 ***
## T                  -0.0311907  0.0004343  -71.81   <2e-16 ***
## Weekend1           -0.9331403  0.0094666  -98.57   <2e-16 ***
## Holiday1           -0.9587444  0.0220115  -43.56   <2e-16 ***
## month               0.0185654  0.0012465   14.89   <2e-16 ***
## factor(Object)G    -0.2385469  0.0157828  -15.11   <2e-16 ***
## factor(Object)L    -1.0107495  0.0157828  -64.04   <2e-16 ***
## factor(Object)Lib  -1.6837879  0.0157828 -106.69   <2e-16 ***
## factor(Object)M    -2.5376071  0.0157828 -160.78   <2e-16 ***
## factor(Object)NVK  -1.4074486  0.0157828  -89.18   <2e-16 ***
## factor(Object)T    -3.0232510  0.0157828 -191.55   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3711 on 7731 degrees of freedom
## Multiple R-squared:  0.9074, Adjusted R-squared:  0.9073
## F-statistic:  7579 on 10 and 7731 DF,  p-value: < 2.2e-16
```

```
# residuals plot
par(mfrow=c(2,2))
plot(fixed.dum)
```

```r
# Fixed effects model is OLS with dummy variable model:
# Each component of the factor variable (Object) is absorbing the effects particular
to each university building

yhat <- fixed.dum$fitted
library(car)
scatterplot(yhat~panel$logcons|panel$Object, boxplots=FALSE, xlab="Cons",
ylab="yhat",smooth=FALSE)
```

```r
# Comparing OLS vs Least squares dummy variable model (LSDV) model

library(apsrtable)
apsrtable(ols, ols1, fixed.dum, model.names = c("OLS", "OLS1", "OLS_DUM"))
## \begin{table}[!ht]
## \caption{}
## \label{}
## \begin{tabular}{ l D{.}{.}{2}D{.}{.}{2}D{.}{.}{2} }
## \hline
##   & \multicolumn{ 1 }{ c }{ OLS } & \multicolumn{ 1 }{ c }{ OLS1 } & \multicolumn{
1 }{ c }{ OLS_DUM } \\ \hline
##  %                  & OLS       & OLS1      & OLS_DUM \\
## (Intercept)        & 6.11 ^*   & 5.68 ^*   & 7.41 ^* \\
##                    & (0.02)    & (0.08)    & (0.01)  \\
## T                  & -0.03 ^* & -0.03 ^* & -0.03 ^*\\
##                    & (0.00)    & (0.00)    & (0.00)  \\
## Weekend1           & -0.93 ^* & -0.94 ^* & -0.93 ^*\\
##                    & (0.03)    & (0.03)    & (0.01)  \\
## Holiday1           & -0.97 ^* & -0.96 ^* & -0.96 ^*\\
##                    & (0.07)    & (0.07)    & (0.02)  \\
## month              &           & 0.01 ^*   & 0.02 ^* \\
##                    &           & (0.00)    & (0.00)  \\
## U                  &           & 0.00 ^*   &         \\
##                    &           & (0.00)    &         \\
## factor(Object)G    &           &           & -0.24 ^*\\
##                    &           &           & (0.02)  \\
## factor(Object)L    &           &           & -1.01 ^*\\
##                    &           &           & (0.02)  \\
## factor(Object)Lib  &           &           & -1.68 ^*\\
##                    &           &           & (0.02)  \\
## factor(Object)M    &           &           & -2.54 ^*\\
##                    &           &           & (0.02)  \\
## factor(Object)NVK  &           &           & -1.41 ^*\\
##                    &           &           & (0.02)  \\
## factor(Object)T    &           &           & -3.02 ^*\\
##                    &           &           & (0.02)   \\
##  $N$               & 7742      & 7742      & 7742    \\
## $R^2$              & 0.19      & 0.19      & 0.91    \\
## adj. $R^2$         & 0.19      & 0.19      & 0.91    \\
## Resid. sd          & 1.10      & 1.10      & 0.37     \\ \hline
##  \multicolumn{4}{l}{\footnotesize{Standard errors in parentheses}}\\
## \multicolumn{4}{l}{\footnotesize{$^*$ indicates significance at $p< 0.05 $}}
## \end{tabular}
##  \end{table}
# The coefficient of xi in OLS indicates how much Y changes when X increases by one
unit.
# The coefficient of xi indicates how much Y changes overtime, controlling by
differences in objects, when X increases by one unit.
# Fixed effects: n entity-specific intercepts (using plm)

# The coeff of xi indicates how much Y changes overtime, on average per school, when X
increases by one unit.

library(plm)
fixed <- plm(logcons ~  T + Weekend + Holiday + month, data=panel, index=c("Object",
"Date"), model="within")
```

```
summary(fixed)
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = logcons ~ T + Weekend + Holiday + month, data = panel,
##     model = "within", index = c("Object", "Date"))
##
## Balanced Panel: n = 7, T = 1106, N = 7742
##
## Residuals:
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -1.77538 -0.22265  0.03112  0.24888  1.47253
##
## Coefficients:
##             Estimate  Std. Error t-value  Pr(>|t|)
## T        -0.03119072  0.00043433 -71.814 < 2.2e-16 ***
## Weekend1 -0.93314027  0.00946662 -98.572 < 2.2e-16 ***
## Holiday1 -0.95874436  0.02201153 -43.556 < 2.2e-16 ***
## month     0.01856543  0.00124653  14.894 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    3244.2
## Residual Sum of Squares: 1065
## R-Squared:      0.67173
## Adj. R-Squared: 0.67131
## F-statistic: 3954.97 on 4 and 7731 DF, p-value: < 2.22e-16
fixed2 <- plm(logcons ~ light + Weekend + Holiday + month + U, data=panel,
index=c("Object", "Date"), model="within")
summary(fixed2)
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = logcons ~ light + Weekend + Holiday + month + U,
##     data = panel, model = "within", index = c("Object", "Date"))
##
## Balanced Panel: n = 7, T = 1106, N = 7742
##
## Residuals:
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -1.93975 -0.22482  0.02497  0.24301  1.50687
##
## Coefficients:
##             Estimate  Std. Error t-value  Pr(>|t|)
## light    -2.21035595  0.04319954 -51.166 < 2.2e-16 ***
## Weekend1 -0.93844955  0.00939874 -99.849 < 2.2e-16 ***
## Holiday1 -0.98435846  0.02182820 -45.096 < 2.2e-16 ***
## month    -0.01380656  0.00122668 -11.255 < 2.2e-16 ***
## U         0.00415767  0.00032818  12.669 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    3244.2
## Residual Sum of Squares: 1046.8
## R-Squared:      0.67732
## Adj. R-Squared: 0.67686
## F-statistic: 3245.12 on 5 and 7730 DF, p-value: < 2.22e-16
#effects for each object:
summary(fixef(fixed))
##     Estimate Std. Error t-value  Pr(>|t|)
## C   7.410431   0.014234  520.62 < 2.2e-16 ***
## G   7.171884   0.014234  503.86 < 2.2e-16 ***
## L   6.399682   0.014234  449.61 < 2.2e-16 ***
## Lib 5.726643   0.014234  402.33 < 2.2e-16 ***
## M   4.872824   0.014234  342.34 < 2.2e-16 ***
## NVK 6.002982   0.014234  421.74 < 2.2e-16 ***
## T   4.387180   0.014234  308.22 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#deviation from overall mean:
summary(fixef(fixed,type="dmean"))
##       Estimate Std. Error   t-value Pr(>|t|)
## C    1.4144844  0.0142339   99.3746   <2e-16 ***
## G    1.1759375  0.0142339   82.6155   <2e-16 ***
## L    0.4037349  0.0142339   28.3644   <2e-16 ***
## Lib -0.2693035  0.0142339  -18.9199   <2e-16 ***
## M   -1.1231227  0.0142339  -78.9050   <2e-16 ***
## NVK  0.0070358  0.0142339    0.4943   0.6211
## T   -1.6087666  0.0142339 -113.0239   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
# Function: Calculates AIC based on an lm or plm object

AIC_adj <- function(fixed){
  # Number of observations
  n.N <- nrow(fixed$model)
  # Residuals vector
  u.hat <- residuals(fixed)
  # Variance estimation
  s.sq  <- log( (sum(u.hat^2)/(n.N)))
  # Number of parameters (incl. constant) + one additional for variance estimation
  p <- length(coef(fixed)) + 1

  # Note: minus sign cancels in log likelihood
  aic <- 2*p  +  n.N * (  log(2*pi) + s.sq  + 1 )

  return(aic)
}
# In the case of the fixed effects regression, we are accounting for fixed effects (or
electricity consuption independent of time), while the second - random effects model -
is accounting for random effects (including time).
# Interpretation of the coefficients is tricky since they include both the within-
entity and between-entity effects, and represent the average effect of X over Y when X
changes across time and between schools by one unit.

library(plm)
random <- plm(logcons ~ T + Weekend + Holiday + month, data=panel, index=c("Object",
"Date"), model="random")
summary(random)
## Oneway (individual) effect Random Effect Model
##    (Swamy-Arora's transformation)
##
## Call:
## plm(formula = logcons ~ T + Weekend + Holiday + month, data = panel,
##     model = "random", index = c("Object", "Date"))
##
## Balanced Panel: n = 7, T = 1106, N = 7742
##
## Effects:
##                  var std.dev share
## idiosyncratic 0.1378  0.3711   0.1
## individual    1.2447  1.1156   0.9
## theta: 0.99
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -1.761230 -0.222695  0.031735  0.248414  1.476573
##
## Coefficients:
##                Estimate  Std. Error z-value  Pr(>|z|)
## (Intercept)  5.99594664  0.42178712  14.216 < 2.2e-16 ***
## T           -0.03119072  0.00043433 -71.814 < 2.2e-16 ***
## Weekend1    -0.93314027  0.00946662 -98.572 < 2.2e-16 ***
## Holiday1    -0.95874436  0.02201153 -43.556 < 2.2e-16 ***
## month        0.01856543  0.00124653  14.894 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Total Sum of Squares:    3245
## Residual Sum of Squares: 1065.8
## R-Squared:      0.67156
## Adj. R-Squared: 0.67139
## Chisq: 15819.9 on 4 DF, p-value: < 2.22e-16
# H0: the null hypothesis is that the preferred model is random effects vs. the
alternative the fixed effects.
# It basically tests whether the unique errors (ui) are correlated with the
regressors, the null hypothesis is they are not.
# If the p-value is significant (for example <0.05) then use fixed effects, if not use
random effects.

phtest(fixed, random)
##
##  Hausman Test
##
## data:  logcons ~ T + Weekend + Holiday + month
## chisq = 3.2557e-11, df = 4, p-value = 1
## alternative hypothesis: one model is inconsistent
# fixed effects is better
library(plm)
# fixed effects model

# Lagrange Multiplier Test - time effects (Breusch-Pagan)
 plmtest(fixed, c("time"), type=("bp"))
##
##  Lagrange Multiplier Test - time effects (Breusch-Pagan) for balanced
##  panels
##
## data:  logcons ~ T + Weekend + Holiday + month
## chisq = 351.83, df = 1, p-value < 2.2e-16
## alternative hypothesis: significant effects
 # p is < 0.05 then use time-fixed effects.

# fixed_time  <- plm(logcons ~ T + Weekend + Holiday + month, data=panel,
index=c("Object", "Date"), model="within", effect="twoways")

# The model with lags of logcons
fixed_lag <- plm(logcons ~ lag(logcons, 1) + lag(logcons, 3) + lag(logcons, 7) +
lag(logcons, 30) + T + Weekend + Holiday + month, data=panel, index=c("Object",
"Date"), model="within")
summary(fixed_lag)
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = logcons ~ lag(logcons, 1) + lag(logcons, 3) + lag(logcons,
##     7) + lag(logcons, 30) + T + Weekend + Holiday + month, data = panel,
##     model = "within", index = c("Object", "Date"))
##
## Balanced Panel: n = 7, T = 1076, N = 7532
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -1.744240 -0.153718 -0.010238  0.147825  1.731501
##
## Coefficients:
##                      Estimate  Std. Error  t-value  Pr(>|t|)
## lag(logcons, 1)    0.18895811  0.00663084  28.4969 < 2.2e-16 ***
## lag(logcons, 3)    0.09931343  0.00608588  16.3187 < 2.2e-16 ***
## lag(logcons, 7)    0.45007476  0.00812510  55.3931 < 2.2e-16 ***
## lag(logcons, 30) -0.07296043  0.00626690 -11.6422 < 2.2e-16 ***
## T                -0.01027281  0.00044229 -23.2264 < 2.2e-16 ***
## Weekend1         -0.50468620  0.01032270 -48.8909 < 2.2e-16 ***
## Holiday1         -0.85043209  0.01677580 -50.6940 < 2.2e-16 ***
## month             0.00375608  0.00097804   3.8404 0.0001238 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    3136.5
```

```
## Residual Sum of Squares: 566.13
## R-Squared:      0.8195
## Adj. R-Squared: 0.81916
## F-statistic: 4266.08 on 8 and 7517 DF, p-value: < 2.22e-16
# Testing for cross-sectional dependence/contemporaneous correlation: using Breusch-
Pagan LM test of independence and Pasaran CD test/

# We accept H0 (no cross-sectional dependence) if p>0.05).

# Breusch-Pagan LM test for cross-sectional dependence in panels
pcdtest(fixed, test = c("lm"))
##
##  Breusch-Pagan LM test for cross-sectional dependence in panels
##
## data:  logcons ~ T + Weekend + Holiday + month
## chisq = 3476.8, df = 21, p-value < 2.2e-16
## alternative hypothesis: cross-sectional dependence
# Pesaran CD test for cross-sectional dependence in panels
pcdtest(fixed, test = c("cd"))
##
##  Pesaran CD test for cross-sectional dependence in panels
##
## data:  logcons ~ T + Weekend + Holiday + month
## z = 42.026, p-value < 2.2e-16
## alternative hypothesis: cross-sectional dependence
# The null hypothesis in the B-P/LM and Pasaran CD tests of independence is that
residuals across entities are not correlated. B-P/LM and Pasaran CD (cross-sectional
dependence) tests are used to test whether the residuals are correlated across
entities*. Cross-sectional dependence can lead to bias in tests results (also called
contemporaneous correlation).

# According to Baltagi, cross-sectional dependence is a problem in macro panels with
long time series. This is not much of a problem in micro panels (few years and large
number of cases).
# Serial correlation tests apply to macro panels with long time series. Not a problem
in micro panels (with very few years). The null is that there is not serial
correlation. (We accept H0 (no serial correlation) if p>0.05).

# Breusch-Godfrey/Wooldridge test for serial correlation in panel models
pbgtest(random)
##
##  Breusch-Godfrey/Wooldridge test for serial correlation in panel models
##
## data:  logcons ~ T + Weekend + Holiday + month
## chisq = 5370.3, df = 1106, p-value < 2.2e-16
## alternative hypothesis: serial correlation in idiosyncratic errors
pbgtest(fixed)
##
##  Breusch-Godfrey/Wooldridge test for serial correlation in panel models
##
## data:  logcons ~ T + Weekend + Holiday + month
## chisq = 5367.8, df = 1106, p-value < 2.2e-16
## alternative hypothesis: serial correlation in idiosyncratic errors
# Breusch-Pagan test

# The null hypothesis for the Breusch-Pagan test is homoskedasticity

library(lmtest)
bptest(random, studentize=F)
##
##  Breusch-Pagan test
##
## data:  random
## BP = 95.723, df = 4, p-value < 2.2e-16
bptest(fixed, studentize=F)
##
##  Breusch-Pagan test
##
## data:  fixed
```

```
## BP = 95.723, df = 4, p-value < 2.2e-16
bptest(fixed_lag, studentize=F)
##
##  Breusch-Pagan test
##
## data:  fixed_lag
## BP = 1448.3, df = 8, p-value < 2.2e-16
# p<0.05 indicates presence of heteroskedasticity

# If hetersokedaticity is detected you can use robust covariance matrix to account for
it.
# The --vcovHC- function estimates three heteroskedasticity-consistent covariance
estimators

# "white1" - for general heteroskedasticity but no serial correlation. Recommended for
random effects.
# "white2" - is "white1" restricted to a common variance within groups. Recommended
for random effects.
# "arellano" - both heteroskedasticity and serial correlation. Recommended for fixed
effects.


# The following options apply:
# HC0 - heteroskedasticity consistent. The default.
# HC1,HC2, HC3 - Recommended for small samples. HC3 gives less weight to influential
observations.
# HC4 - small samples with influential observations
# HAC - heteroskedasticity and autocorrelation consistent (type ?vcovHAC for more
details)

# Controlling for heteroskedasticity: Fixed effects

coeftest(fixed) # Original coefficients
##
## t test of coefficients:
##
##              Estimate  Std. Error t value  Pr(>|t|)
## T        -0.03119072  0.00043433 -71.814 < 2.2e-16 ***
## Weekend1 -0.93314027  0.00946662 -98.572 < 2.2e-16 ***
## Holiday1 -0.95874436  0.02201153 -43.556 < 2.2e-16 ***
## month     0.01856543  0.00124653  14.894 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
coeftest(fixed, vcovHC) # Heteroskedasticity consistent coefficients
##
## t test of coefficients:
##
##             Estimate Std. Error t value  Pr(>|t|)
## T        -0.0311907  0.0040267 -7.7460 1.069e-14 ***
## Weekend1 -0.9331403  0.1372334 -6.7997 1.127e-11 ***
## Holiday1 -0.9587444  0.1385408 -6.9203 4.867e-12 ***
## month     0.0185654  0.0026515  7.0018 2.739e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
coeftest(fixed_lag) # Original coefficients
##
## t test of coefficients:
##
##                      Estimate  Std. Error  t value  Pr(>|t|)
## lag(logcons, 1)   0.18895811  0.00663084  28.4969 < 2.2e-16 ***
## lag(logcons, 3)   0.09931343  0.00608588  16.3187 < 2.2e-16 ***
## lag(logcons, 7)   0.45007476  0.00812510  55.3931 < 2.2e-16 ***
## lag(logcons, 30) -0.07296043  0.00626690 -11.6422 < 2.2e-16 ***
## T                -0.01027281  0.00044229 -23.2264 < 2.2e-16 ***
## Weekend1         -0.50468620  0.01032270 -48.8909 < 2.2e-16 ***
## Holiday1         -0.85043209  0.01677580 -50.6940 < 2.2e-16 ***
## month             0.00375608  0.00097804   3.8404 0.0001238 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
coeftest(fixed_lag, vcovHC) # Heteroskedasticity consistent coefficients
##
## t test of coefficients:
##
##                     Estimate  Std. Error  t value  Pr(>|t|)
## lag(logcons, 1)   0.18895811  0.01759575  10.7388 < 2.2e-16 ***
## lag(logcons, 3)   0.09931343  0.01707851   5.8151 6.307e-09 ***
## lag(logcons, 7)   0.45007476  0.03199564  14.0668 < 2.2e-16 ***
## lag(logcons, 30) -0.07296043  0.00656805 -11.1084 < 2.2e-16 ***
## T                -0.01027281  0.00170268  -6.0333 1.682e-09 ***
## Weekend1         -0.50468620  0.10069076  -5.0122 5.503e-07 ***
## Holiday1         -0.85043209  0.12673170  -6.7105 2.081e-11 ***
## month             0.00375608  0.00078411   4.7902 1.698e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
# calculate fitted values for model fixed_lag
fitted_lag <- as.numeric(fixed_lag$model[[1]] - fixed_lag$residuals)
plot(fitted_lag, col=3)
```

```
line(panel$logcons)
##
## Call:
## line(panel$logcons)
##
## Coefficients:
## [1]   7.0078942  -0.0003792
# calculate fitted values for model fixed
fitted_fix <- as.numeric(fixed$model[[1]] - fixed$residuals)
plot(fitted_fix, col=2)
```

```
line(panel$logcons)
##
## Call:
## line(panel$logcons)
##
## Coefficients:
## [1]   7.0078942  -0.0003792
# plot fitted values
library(car)
yhat <- fixed.dum$fitted
scatterplot(yhat~panel$T|panel$Object, boxplots=FALSE, xlab="Cons",
ylab="yhat",smooth=FALSE)
abline(lm(panel$logcons~panel$T),lwd=3, col="red")
```

```
# make prediction
plot(panel$T, panel$logcons, pch=19, xlab="temperature", ylab="consumption")
abline(plm(logcons ~ T + Weekend + Holiday + month, data=panel, index=c("Object",
"Date"), model="within"),lwd=3, col="red")
## Warning in abline(plm(logcons ~ T + Weekend + Holiday + month, data = panel, :
## only using the first two of 4 regression coefficients
```

```
plot(panel$light, panel$logcons, pch=19, xlab="light", ylab="consumption")
abline(plm(logcons ~ light + Weekend + Holiday, data=panel, index=c("Object", "Date"),
model="within"),lwd=3, col="red")
## Warning in abline(plm(logcons ~ light + Weekend + Holiday, data = panel, : only
## using the first two of 3 regression coefficients
```

```
# plot fitted vs actual for fixed.dum
yhat <- fixed.dum$fitted
ggplot() + geom_point(data=panel, aes(x=T, y=logcons)) +
```

```
    geom_line(data=panel,aes(x=T, y= yhat), color = "red")
```

```
    #facet_wrap(Holiday ~ Weekend)

# New data for Forecast el consumption gor obj. G working day
newdataG = data_frame(T=2, Weekend='0', Holiday='0', month=12, Object='G')
## Warning: `data_frame()` is deprecated as of tibble 1.1.0.
## Please use `tibble()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.
# New data for Forecast el consumption gor obj. G weekend
newdataG_W = data_frame(T=2, Weekend='1', Holiday='0', month=12, Object='G')

# New data for Forecast el consumption gor obj. C working day
newdataC = data_frame(T=2, Weekend='0', Holiday='0', month=12, Object='C')
# New data for Forecast el consumption gor obj. C weekend
newdataC_W = data_frame(T=2, Weekend='1', Holiday='0', month=12, Object='C')

# Forecasts
exp(predict(fixed.dum, newdataG, interval = "confidence"))
##        fit      lwr      upr
## 1 1528.876 1487.422 1571.485
exp(predict(fixed.dum, newdataG_W, interval = "confidence"))
##        fit      lwr      upr
## 1 601.3323 583.6025 619.6007
exp(predict(fixed.dum, newdataG, interval = "prediction"))
##        fit      lwr      upr
## 1 1528.876 738.2018 3166.425
exp(predict(fixed.dum, newdataG_W, interval = "prediction"))
##        fit      lwr      upr
## 1 601.3323 290.3191 1245.528
exp(predict(fixed.dum, newdataC, interval = "prediction"))
##       fit      lwr      upr
## 1 1940.76 937.0758 4019.471
exp(predict(fixed.dum, newdataC_W, interval = "prediction"))
##        fit      lwr      upr
## 1 763.3331 368.5321 1581.077
exp(predict(fixed.dum, newdataC_W, interval = "confidence"))
##        fit      lwr      upr
## 1 763.3331 740.8269 786.5231
# Choose the best model

cons_cg <- read.csv("el_cons_C_G_2015_20.csv", dec = ".", sep = ",")
cons_cg <- cons_cg[1:2131,]

library(forecast)
## Warning: package 'forecast' was built under R version 4.0.3
# Convert data to ts object
dat_ts <- ts(cons_cg[,-1], start = c(2015, 1, 1), frequency = 365.25)

library(forecast)
train <- window(dat_ts, start=c(2015,1), end=c(2020,274))
test <- window(dat_ts, start=c(2020,274))

x_g <- train[,"Conc_G"]
x_c <- train[,"Cons_C"]
x_g_t <- test[,"Conc_G"]
x_c_t <- test[,"Cons_C"]

par(mfrow=c(1,2))
plot.ts(dat_ts[,"Conc_G"], ann=FALSE)
plot.ts(dat_ts[,"Cons_C"], ann=FALSE)
```

```
# Create a QQ-plot for the 3rd column of the data set 'dat' - 'Dep_Rate' to check if
it is normally distributed
# Function par(mfrow=c(1,2)) specifies location of two charts in 1 row and 2 columns
```

```
par(mfrow=c(2,2))
qqnorm(dat_ts[ ,1])
qqline(dat_ts[ ,1])

qqnorm(dat_ts[ ,2])
qqline(dat_ts[ ,2])
# Create a QQ-plot for the first differences of the logarithm of 'Dep_Rate'
(TRANSFORMED TS)

# Operator 'log' (logarithmization) is used as transformation to reduce the data
variability
# Operator 'diff' is used to transform data to stationary view - diff(y(i))=y(t-1)-
y(t)
par(mfrow=c(2,2))
```

```
qqnorm(diff(log(dat_ts[ ,1])))
qqline(diff(log(dat_ts[ ,1])))

qqnorm(diff(log(dat_ts[ ,2])))
qqline(diff(log(dat_ts[ ,2])))
```

```
# El consumption and temperature
autoplot(dat_ts[,c("Conc_G","T")], facets=TRUE) +
  xlab("") + ylab("") +
  ggtitle("Daily electricity demand: G")
```

```
# Seasonal plot by years
ggseasonplot(dat_ts[ ,1], year.labels=TRUE, year.labels.left=TRUE) +
  ylab("$ million") +
  ggtitle("Seasonal plot by years")
```

```
# Autocorrelation plots
ggAcf(dat_ts[ ,1])
```

```
ggAcf(dat_ts[ ,1], lag=28) # for lags = 7*4= 28
```

```
# Series decomposition
dat_ts[ ,1] %>% decompose(type="multiplicative") %>%
  autoplot() + xlab("Year") +
  ggtitle("Classical multiplicative decomposition
    of electrical consumption for obj. G")
```

```
#Model using different method
#arima, expo smooth, theta, random walk, structural time series
models<-list(
#arima
mod_arima <- forecast(auto.arima(x_g, ic='aicc', stepwise=FALSE, xreg=train[,c("T",
"U", "Weekend", "Holiday", "light_av",  "month")]), h=31, xreg=test[,c("T",  "U",
"Weekend", "Holiday", "light_av",  "month")]),
#exp smoothing
mod_exponential <- forecast(ets(x_g, ic='aicc', restrict=FALSE), h=31),
#neural networks
mod_neural <- forecast(nnetar(x_g, p=12, size=25), h=31),
#TBATS model (Exponential smoothing state space model with Box-Cox transformation,
ARMA errors, Trend and Seasonal components)
mod_tbats <- forecast(tbats(x_g, ic='aicc', seasonal.periods=7), h=31),
#random walk
mod_rw <-rwf(x_g, h=31)
```

```
)
## Warning in ets(x_g, ic = "aicc", restrict = FALSE): I can't handle data with
## frequency greater than 24. Seasonality will be ignored. Try stlf() if you need
## seasonal forecasts.
##Compare the training set forecast with test set
par(mfrow=c(3, 2))
for (f in models){
    plot(f)
    lines(test[,"Conc_G"],col='red')
}

##To see its accuracy on its Test set,
#as training set would be "accurate" in the first place
acc.test<-lapply(models, function(f){
    accuracy(f, test[,"Conc_G"])[2,]
})
acc.test <- Reduce(rbind, acc.test)
row.names(acc.test)<-c("arima","expsmooth","neural","tbats","randomwalk")
acc.test <- acc.test[order(acc.test[,'MASE']),]
acc.test
##                       ME      RMSE       MAE         MPE      MAPE      MASE
## tbats         -5.457171 113.9171  75.41876  -1.509800 13.26622 0.2042835
## arima        -58.338166 141.0247 125.26701  -2.789818 20.24544 0.3393052
## expsmooth    -19.844797 209.5170 183.00400 -17.139237 36.67837 0.4956949
## neural      -210.431063 293.0116 232.17096 -29.015580 34.47153 0.6288712
## randomwalk -234.755161 314.0280 236.26484 -53.411569 53.57660 0.6399601
##                    ACF1 Theil's U
## tbats       0.28933120 0.3560603
## arima       0.20063025 0.4066726
## expsmooth   0.06955319 0.7272666
## neural      0.48086727 0.8059673
## randomwalk  0.06955319 1.0711524
##Look at training set to see if there are overfitting of the forecasting
##on training set
acc.train<-lapply(models, function(f){
    accuracy(f, test[,"Conc_G"])[1,]
})
acc.train <- Reduce(rbind, acc.train)
row.names(acc.train)<-c("arima","expsmooth","neural","tbats","randomwalk")
acc.train <- acc.train[order(acc.train[,'MASE']),]
acc.train
##                        ME      RMSE       MAE         MPE      MAPE      MASE
## neural       0.006374821 145.6561  96.98256  -5.101076 14.20115 0.2626924
## arima       -0.843845227 191.3023 142.92861  -2.355141 20.14721 0.3871444
## tbats       27.297382472 241.1631 155.07430  -4.756027 19.83911 0.4200429
## randomwalk   0.009256789 468.9530 288.78387 -16.649778 39.46402 0.7822161
## expsmooth   -1.024617317 443.1668 360.92869 -31.148058 57.87045 0.9776316
##                     ACF1 Theil's U
## neural      -0.008169331        NA
## arima        0.072516222        NA
## tbats        0.038867952        NA
## randomwalk -0.022793015        NA
## expsmooth    0.384039401        NA
# check the order of the estimated ARIMA MODEL
arimaorder(auto.arima(x_g, ic='aicc', stepwise=FALSE, xreg=train[,c("T",  "U",
"Weekend", "Holiday", "light_av",  "month")]))
## p d q
## 2 1 3
```