Journal of Infrastructure, Policy and Development 2025, 9(2), 10621.
https://doi.org/10.24294/jipd10621

Article

Comparative modeling of household electricity consumption in France: Insights from path analysis and classical models

Seyid Abdellahi Ebnou Abdem [1,*], Mariem Bounabi [1], El Bachir Diop [1], Rida Azmi [1], Mohammed Hlal [1], Meriem Adraoui [1], Imane Serbouti [1], Jérôme Chenal [1,2]

[1] Center of Urban Systems (CUS), University Mohammed VI Polytechnic (UM6P), Ben Guerir 43150, Morocco
[2] Urban and Regional Planning Community (CEAT), Ecole Polytechnique Fédérale de Lausanne (EPFL), 1266 Lausanne, Switzerland
* **Corresponding author:** Seyid Abdellahi Ebnou Abdem, seyid.ebnouabdem@um6p.ma

CITATION

Abdem SAE, Bounabi M, Diop EB. (2025). Comparative modeling of household electricity consumption in France: Insights from path analysis and classical models. Journal of Infrastructure, Policy and Development. 9(2): 10621. https://doi.org/10.24294/jipd10621

ARTICLE INFO

Received: 27 November 2024
Accepted: 10 December 2024
Available online: 8 February 2025

Abstract: Electricity consumption in Europe has risen significantly in recent years, with households being the largest consumers of final electricity. Managing and reducing residential power consumption is critical for achieving efficient and sustainable energy management, conserving financial resources, and mitigating environmental effects. Many studies have used statistical models such as linear, multinomial, ridge, polynomial, and LASSO regression to examine and understand the determinants of residential energy consumption. However, these models are limited to capturing only direct effects among the determinants of household energy consumption. This study addresses these limitations by applying a path analysis model that captures the direct and indirect effects. Numerical and theoretical comparisons that demonstrate its advantages and efficiency are also given. The results show that Sub-metering components associated with specific uses, like cooking or water heating, have significant indirect impacts on global intensity through active power and that the voltage affects negatively the global power (active and reactive) due to the physical and behavioral mechanisms. Our findings provide an in-depth understanding of household electricity power consumption. This will improve forecasting and enable real-time energy management tools, extending to the design of precise energy efficiency policies to achieve SDG 7's objectives.

Keywords: household power consumption; regression models; residential electricity modeling; path analysis model

1. Introduction

Sustainable Development Goal 7 (SDG 7, Affordable and Clean Energy) is the key to achieving climate goals, and constructing effective policies has an immense role in fostering these actions. Residential electricity consumption is a major component of SDG 7, and its influence is predominantly on all three components of SDG 7 [1]. Electricity plays a main role in modern economies and in the increasing demand for energy services [2]. The electricity demand is increasing internationally in all sectors due to factors including increased family incomes, technological advancements and the Internet of Things (IoTs), electrification of the transportation system, the replacement of non-electric heaters with electric ones, and increased usage of air conditioning [2]. The residential sector is always the world's second or third largest final electricity consumer [2].

Electricity consumption in Europe has been on the rise in recent years.

Consequently, the world's energy shortage is getting worse and worse. Growth in global electricity demand in 2024 and 2025 is set to be among the fastest in the past two decades [3]. Based on the report of IEA's Electricity Mid-Year Update, electricity demand is forecast to grow by around 4% in 2024, up from 2.5% in 2023, and this is the highest annual growth rate since 2007, excluding the exceptional rebounds seen in the wake of the global financial crisis and the Covid-19 pandemic [3]. In European countries, including France, the residential sector is one of the largest consumers of electricity, accounting for about a third of the total final electricity demand [4]. Besides, many European countries, including France, have launched an urgency to reduce greenhouse gas emissions [5]. In 2023, 65% of the electricity produced in France came from nuclear power, and solar energy sources accounted for 14% [6].

According to the International Energy Agency's 2013 reference scenario, by 2040, 14% of worldwide energy consumption will come from households, a 57% increase compared with the 2010 rate [7]. The residential sector is thus responsible for a large proportion of energy consumption [7]. In France, the residential sector is the second largest source of consumption, at 46 million metric tons of oil equivalent (Mtoe) in 2012, just behind the transport sector [7]. Apart from the fact that energy efficiency in buildings represents the largest source of energy consumption, it also largely reflects household energy practices. This explained the increasing and interrelated interest of researchers and governments in understanding the determinants of domestic energy use to develop measures to rationalize consumption [7–9].

Furthermore, according to the "Energy Balance of France" published by the Department of Observation and Statistics (SOeS), the residential sector remains the second final energy consumer after transport (49 Mteo), with 30% of final energy consumption and nearly 20% of greenhouse gas emissions [9].

The significant increase in household energy consumption, frequent electricity shortages and blackouts, and rising electricity prices in the residential context [4,7–11]. Given these challenges, it is imperative to adopt efficient energy usage strategies [4,9–11]. To adopt this efficient and optimized usage, we need a significant reduction in residential energy demand. Despite all this, compared to the three end-use sectors (transportation, industrial, etc.), the residential sector is largely understudied [7,9,10]. This scarcity of studies has driven many researchers and policymakers to try to better understand the determinants of energy consumption and to identify conservation strategies [4,8–11].

In the past few years, many empirical studies have used statistical models [4, 7, 9] and machine learning (ML) approaches [10] to examine and understand the determinants of residential energy consumption. However, modeling household energy consumption is a complex issue because it is sorely linked to the multitude of interrelated factors, such as technical attributes of the buildings, household characteristics, and behavior [9], energy systems, climate, equipment, and socio-cultural factors (e.g., household size and composition, greying of society) [8].

Many recent studies have been conducted to address this issue: [12] used a quantile regression model to examine predictors of household energy consumption among single-family residences; [9] used a multivariate regression analysis (Specifically, log-linear models based on Ordinary Least Square (OLS)) to tease out the impacts of various factors on the domestic energy consumption in France; and [10] offers an extendable experimental analytical framework to modeling household electric power consumption in France by comparing multinomial, ridge, polynomial, and Least Absolute Shrinkage and Selection Operator (Lasso) regression models.

Unfortunately, these models are limited in their ability to capture complex indirect effects [8,11] among the determinants of household energy consumption. In this study, we have proposed a Path Analysis Model (PAM) that overcomes these limitations by capturing and understanding the components of the direct and indirect effects on household energy consumption. This model is based on a dataset of over 2 million users spanning four years in France. This study examined the key characteristics of power consumption, analyzed energy consumption patterns, identified unusual behaviors about household characteristics, and offered insights into the key determinants of residential energy consumption to inform energy policies. The findings of the proposed model provide deeper insight into the effect of building and occupant attributes on electricity use.

In addition to this introduction (Section 1), the rest of this paper is structured as follows: Section 2 details the data source and description, our hypothesis, and our proposed model. Section 3 presents the findings and discussion. Finally, Section 4 offers the main conclusions, perspectives, and policy implications.

2. Methodology

This section describes the scope and context of this study, the data source, and the description. In addition, it presents our hypothesis and our proposed model.

2.1. Data description

The dataset used in this research comprises 2,075,259 observations of household electric power consumption, collected at a one-minute interval over a period of four years (2006 to 2010) from a residence in *Sceaux* (7 km of Paris, France) [10,13–15]. This dataset, licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, is publicly available, and used in the Centre for Machine Learning and Intelligent Systems [10,13–15]. The dataset contains seven key variables, which are [10,13–15] :

- **Global active power**: Global quantity of electrical energy consumed in the household.
- **Global reactive power**: Is the global quantity of electrical power that oscillates between the load and the source without performing any useful work.
- **Volatge**: Minute-averaged voltage.
- **Global intensity**: Household global minute-averaged current intensity.
- **Sub metering 1**: Energy consumption of the first sub-meter, typically associated with laundry room appliances (washing machine, tumble dryer, refrigerator, and lights).
- **Sub metering 2**: Energy consumption of the second sub-meter, often linked to kitchen appliances.
- **Sub metering 3**: Energy consumption of the third sub-meter, typically associated with water heating and air conditioning.

To ensure the integrity and accuracy of the analysis, we performed an in-depth data cleaning to keep the analysis robust, which included processing missing values (that represent only 1.25% [13–15]) but also checking for outliers or inconsistencies that may have been biasing results. This step was done using specialized libraries in Python and *R*. **Table 1** below contains the statistical summary of these variables.

After finishing the data sourcing, cleaning, and presentation of descriptive statistics for each variable, we move to an exploratory analysis focused on examining the correlations among these variables. The next part 2.2 is dedicated to this analysis.

Table 1. Summary statistics of the variables.

Statistic	Voltage	Sub Metering1	Sub Metering2	Sub Metering3	Global Active Power	Global Reactive Power	Global Intensity
Min	−5.0144	−0.1863	−0.2322	−0.7227	−0.9138	−1.0801	−0.9485
1st Qu.	−0.6058	−0.1863	−0.2322	−0.7227	−0.7261	−1.0801	−0.7385
Median	0.0762	−0.1863	−0.2322	−0.7227	−0.4942	−0.1851	−0.4866
Mean	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3rd Qu.	0.6638	−0.1863	−0.0748	1.3478	0.3858	0.6185	0.3531
Max	3.7084	12.4696	12.0460	3.0530	8.4652	11.6133	8.7501

2.2. Exploratory data analysis and hypotheses

According to **Table 1**, all variables are continuous. Hence, to explore the interrelationships and potential causal connections among these variables. We standardized these variables and computed the Pearson correlation matrix using respectively the scale, and cor functions in *R*.

Table 2 below presents the first 6 lines of the table extracted after standardization.

Table 2. Extract of the standardized dataset showing six observation records.

Observation	Global Active Power	Global Reactive Power	Voltage	Global Intensity	Sub Metering1	Sub Metering2	Sub Metering3
1	2.7514	2.7371	−1.5588	2.8722	−0.1863	−0.0748	1.3478
2	3.7642	2.9015	−1.9272	3.8378	−0.1863	−0.0748	1.2260
3	3.7766	3.4676	−2.0307	3.8378	−0.1863	0.0826	1.3478
4	3.7890	3.5042	−1.8937	3.8378	−0.1863	−0.0748	1.3478
5	2.2644	3.7416	−1.3030	2.3264	−0.1863	−0.0748	1.3478
6	2.1352	3.6868	−1.5040	2.1584	−0.1863	0.0826	1.3478

Table 3 below presents this correlation matrix.

Table 3. Extract of the standardized dataset showing six observation records.

	Voltage	Sub Metering1	Sub Metering2	Sub Metering3	Global Active Power	Global Reactive Power	Global Intensity
Voltage	1.0000	−0.2038	−0.1789	−0.2849	−0.4010	−0.1206	−0.4113
SubMetering1	−0.2038	1.0000	0.0608	0.1156	0.4741	0.1377	0.4782
SubMetering2	−0.1789	0.0608	1.0000	0.1065	0.4567	0.1505	0.4622
SubMetering3	−0.2849	0.1156	0.1065	1.0000	0.6241	0.0763	0.6132
Global Active Power	−0.4010	0.4741	0.4567	0.6241	1.0000	0.2548	0.9990
Global Reactive Power	−0.1206	0.1377	0.1505	0.0763	0.2548	1.0000	0.2718
Global Intensity	−0.4113	0.4782	0.4622	0.6132	0.9990	0.2718	1.0000

Table 3 shows the correlations between Global Active Power and (Sub Metering1, Sub Metering2, Sub Metering3, and Voltage), Global Reactive Power with (Voltage and Global Active Power), and Global Intensity with (Global Active Power and Global Reactive Power). These correlations indicate potential associations and suggest possible causal effects. To test the significance of these correlations, we compute the

Test for Significance of Pearson's Correlation Coefficient using cor.test function in *R* software [16–18]. **Tables 4–6** below present the results of this test.

Table 4. Test for significance of pearson's correlation coefficients.

Voltage	**Correlation with Global Active Power**	***P*-value**	**Significance**
Voltage	-0.4	$< 2.2 \times 10^{-16}$	Significant
SubMetering1	0.48	$< 2.2 \times 10^{-16}$	Significant
SubMetering2	0.46	$< 2.2 \times 10^{-16}$	Significant
SubMetering3	0.62	$< 2.2 \times 10^{-16}$	Significant

Table 5. Test for significance of pearson's correlation coefficients.

Voltage	**Correlation with Global Active Power**	***P*-value**	**Significance**
Voltage	-0.12	$< 2.2 \times 10^{-16}$	Significant
Global Active Power	0.25	$< 2.2 \times 10^{-16}$	Significant

Table 6. Test for significance of pearson's correlation coefficients

Voltage	**Correlation with Global Intensity**	***P*-value**	**Significance**
Global Reactive Power	0.27	$< 2.2 \times 10^{-16}$	Significant
Global Active Power	0.99	$< 2.2 \times 10^{-16}$	Significant

Tables 4–6 show that all correlations are significant (< 0.05). Hence, based on the significance of these correlations and the existing studies in energy consumption and power demand management [2, 10, 19–22] the following hypotheses are proposed:

1) $\mathbf{H}_1$: Voltage, Sub Metering 1, Sub Metering 2, and Sub Metering 3 positively influence Global Active Power. This hypothesis indicates a direct impact of these sub-metered consumption metrics on overall power usage.
2) $\mathbf{H}_2$: Voltage and Global Active Power positively influence Global Reactive Power. $\mathbf{H}_2$ suggests an interaction (causal effect) between power demands and reactive power management.
3) $\mathbf{H}_3$: Global Active Power and Global Reactive Power positively influence Global Intensity. $\mathbf{H}_3$ indicates that increases in overall and reactive power demand could drive intensity variations.
4) $\mathbf{H}_4$: Sub Metering 1, Sub Metering 2, and Sub Metering 3 indirectly and positively influence Global Reactive Power, potentially through their effects on Global Active Power.
5) $\mathbf{H}_5$: Sub Metering 1, Sub Metering 2, and Sub Metering 3 indirectly and positively influence Global Intensity, mediated through their impact on both Global Active Power and Global Reactive Power.

Each hypothesis reflects and suggests meaningful potential causal effects (direct[1], and indirect[2]) or associative relationships based on the physical interactions expected between these variables, which we propose to confirm or reject by a conceptual model.

For this purpose, PAM is chosen as a conceptual model to test and confirm these hypotheses, especially when both direct and indirect effects are relevant [16, 23–29]. Since PAM can deal with complex relations between more than two variables when both direct and indirect effects are of interest, Consequently, we employed it [16, 23–29]. Other models (Multiple Linear Regression, Multinomial Regression, Ridge Regression, LASSO Regression, Polynomial Regression) [10,19] were considered, but they were later deemed less appropriate for our aims based on their inability to model causal and mediating relationships. A detailed analysis of these alternative models is shown below, which roughly explains why they were not selected.

Multiple Linear Regression (MLR): is a statistical method for estimating the value of a dependent numerical variable using one or more predictor (independent) variables by assuming that the dependent variable is linearly related to the independent variables [16].

In this study, MLR was also considered to examine the relationships between variables proposed in the hypothesis. However, it was identified as inappropriate for our conceptual model given its inability to handle highly complex dependencies where variables function as dependent and independent of other variables in these hypotheses. For instance, in $\mathbf{H}_1$ Global Active Power is a dependent variable that is affected by (Voltage, Sub Metering 1, Sub Metering 2, and Sub Metering 3). In contrast, in $\mathbf{H}_2$ and $\mathbf{H}_3$, Global Active Power itself serves as an independent (predictor) variable affecting both Global Reactive Power and Global Intensity, respectively. Hence Global Active Power is a dependent (outcome) in one hypothesis (i.e. $\mathbf{H}_1$), and a predictor independent in other hypotheses (i.e. $\mathbf{H}_2$, and $\mathbf{H}_3$).

MLR is effective for situations when there is only one dependent variable and multiple predictors influencing it [16,23–25,28,29]. Yet it does not have the ability to model dependencies, where a variable plays the role of dependent in one context and independent in another [16,23,24,26–29]. Therefore, the flexibility afforded by MLR would not enable the exploration of the causal pathways and mediating effects in our hypothesis.

- Multinomial Regression: is a statistical method used to model the relationship between a binary or categorical (polytomous) dependent variable with more than two categories and a set of independent variables (continuous or categorical) [31, 32]. It is an extension of the logistic regression, in which a dependent variable has only a binary choice (e.g., presence/absence of a characteristic), the dependent variable in a multinomial logistic regression model can have more than two choices that are coded categorically, and one of the categories is taken as the reference category [31,32]. Although it is powerful in scenarios where we need to understand the influence of predictors on binary or categorical outcomes, it lacks the flexibility to capture and interpret indirect effects and mediating relationships, which are essential for the complex causal structures in our hypotheses where all variables are continuous (See **Table 1**). This limitation makes it unsuitable for our study.
- Ridge Regression (RR): is an approach for estimating the coefficients of MLR in scenarios where the independent variables are highly correlated [33–35]. It is an extension of ordinary least-squares regression that introduces a regularization term to handle multicollinearity and prevent overfitting [33, 34, 36]. Although RR is an extension of MLR that addresses issues of multicollinearity (when independent variables are highly correlated), it is limited by the obligation to model a single dependent variable and cannot be used to model complex relationships where a variable may be both dependent and independent [34,35]. Hence, it cannot accommodate the hypotheses of complex relationships, where intermediate variables may convey indirect effects.
- LASSO Regression: is a regularization method that overcomes the limitations

of linear regression concerning the instability of estimation and unreliability of prediction in a high-dimensional context [37]. The main advantage of LASSO regression lies in its ability to perform variable selection, which can prove invaluable in the presence of a large number of variables [37]. However, like RR, it is constrained nevertheless by its incapacity to manage indirect effects and causal linkages inside a model. Lasso does not facilitate the simultaneous estimation of links among several variables in a causal network; instead, it concentrates on variable selection and regularization.

- Polynomial Regression: Polynomial regression is a statistical model to study the non-linear relationships between a dependent variable and a set of predictor variables by adding polynomial terms of predictor variables [38]. The core limitations of polynomial regression in our context are: Firstly, it enables only modeling the non-linear relationships; this is not the case for us, as all the causal relationships we have are linear. Secondly, it cannot differentiate between direct and indirect effects, as it focuses on fitting a specific non-linear relationship rather than exploring the complex causal structures.

After justifying the choice of PAM as a conceptual model, the next part 2.3 presented this model's basic definitions, concepts, and assumptions.

2.3. Proposed conceptual model: Path analysis

PAM is a set of statistical techniques used to assess the causal relationships between observed variables [30,39–43]. In other words, PAM is an advanced statistical method for identifying and examining both the direct and indirect causal relationships among a set of exogenous variables (ξ)[3] (independence, predictor, input) [11] and endogenous variables (η)[4] (dependence, output) [11,30]. PAM can be considered an extension of multiple regression models in the sense that it allows several variables to be dependent. Unlike multiple regression, where only one variable is dependent, PAM permits variables to serve as both causes and effects [30,39,43]. **Figure 1** below, which illustrates the limitations of multiple regression models when dealing with multiple intermediate dependent variables or situations where variables form a causal chain (e.g., Z1 depends on Z, which in turn depends on Y). PAM overcomes these limitations by examining both direct and indirect (intermediate) effects [11,30,41,43].

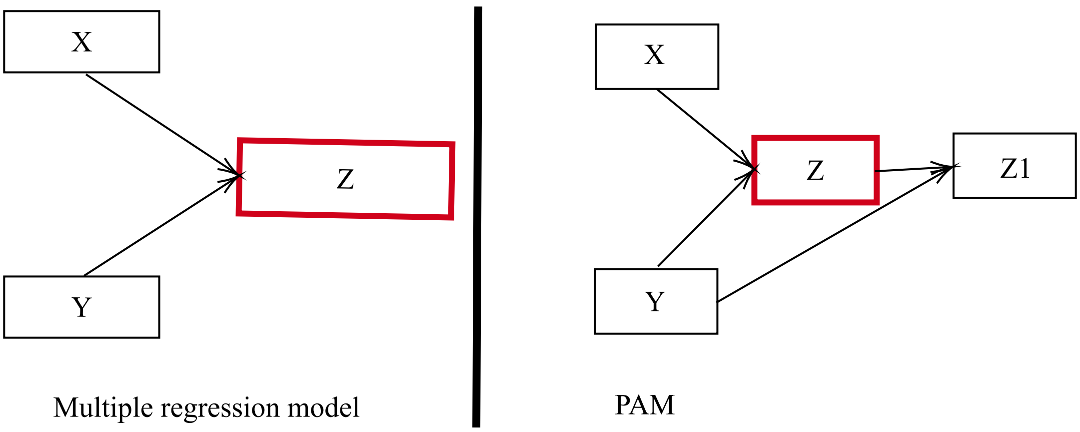

Figure 1. Multiple regression models vs PAM.

On the other hand, PAM is a special case of Structural Equation Modeling (SEM) which is a set of related statistical techniques used to evaluate the fit of a hypothesized causal model with available data, the differentiation of PAM from other SEM models is that it concerns the directs and indirect effects only between the observed variables [11, 30]. PAM was first developed in the first half of the 20th century by Sewall Wright [11, 30]. In recent years, PAM has had many applications in the energy sector, such as in analyzing household energy consumption factors, the energy supply chain, CO2 emission intensity factors, emissions, consumer intentions, and renewable energy [11,44–48]. However, PAM is rarely used to study household electrical power consumption. Most studies of household energy consumption focus on econometric methods, machine learning models, and regression analysis to predict electrical power consumption [10, 11, 19–21, 49]. A PAM can be expressed algebraically as follows [30]:

$$\begin{cases} \eta_1 = \gamma_{11}\xi_1 + \ldots + \gamma_{1q}\xi_q + \zeta_1 \\ \vdots \\ \eta_j = \gamma_{j1}\xi_1 + \ldots + \gamma_{jq}\xi_q + \beta_{j1}\eta_1 + \ldots + \beta_{j(j-1)}\eta_{j-1} + \zeta_j \\ \vdots \\ \eta_p = \gamma_{p1}\xi_1 + \ldots + \gamma_{pq}\xi_q + \beta_{p1}\eta_1 + \ldots + \beta_{p(p-1)}\eta_{p-1} + \zeta_p \end{cases}$$

where q and p are respectively the number of exogenous and endogenous variables, and $\gamma_{i,j}$, and $\beta_{i,j}$ are respectively the coefficients linking endogenous variables to exogenous variables, and the coefficients that relate the endogenous variables between them. In addition, ζ_j where $j = 1 \ldots p$ is the disturbance (error) [30]. The model parameters are estimated by minimizing a specific criterion such as Unweighted Least Square, Maximum likelihood, and Generalized Least Square [30,50].

Figure 2 below depicts an example of PAM with two exogenous variables and two endogenous variables [30].

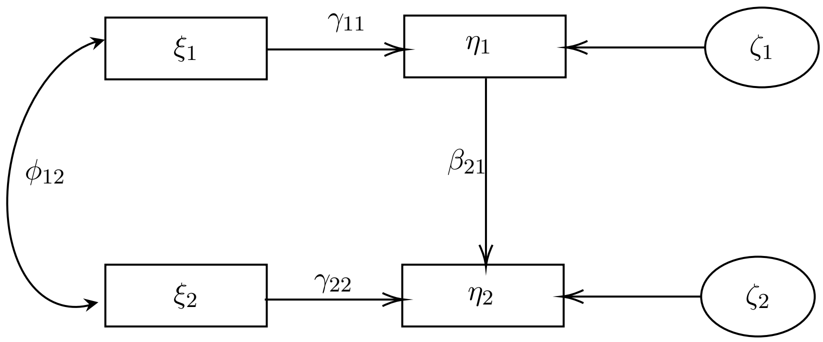

Figure 2. The path diagram for PAM with two exogenous variables, and two endogenous variables [30].

After outlining the general structure and framework of PAM, we now present PAM applied to predict household energy consumption below:

$$\begin{cases} \text{Global Active Power} = \gamma_{11} \times \text{Voltage} + \gamma_{12} \times \text{Sub Metering1} + \gamma_{13} \times \text{Sub Metering2} \\ +\gamma_{14} \times \text{Sub Metering3} + \zeta_1 \\ \\ \text{Global Reactive Power} = \gamma_{21} \times \text{Voltage} + \beta_{21} \times \text{Global Active Power} + \zeta_2 \\ \\ \text{Global Intensity} = \beta_{31} \times \text{Global Active Power} + \beta_{32} \times \text{Global Reactive Power} + \zeta_3 \end{cases} \tag{1}$$

In this model presented in (Equation (1)), Voltage, Sub Metering 1, Sub Metering 2, and Sub Metering 3 are exogenous variables that directly and indirectly influence Global Active Power, Global Reactive Power, and Global Intensity. Global Active

Power is an endogenous variable impacted by all exogenous variables (Voltage, Sub Metering 1, Sub Metering 2, and Sub Metering 3). Global Reactive Power, another endogenous variable, is influenced both directly by the endogenous variable Global Active Power and by the exogenous variable Voltage. Finally, Global Intensity is the last endogenous variable, which is directly influenced by the two endogenous variables (Global Active Power, and Global Reactive Power).

Numerous computational software, such as SPSS, AMOS, LISREL, MPLUS, *R*-Studio, and Python, can be used to fit the model defined in (Equation (1)). We will use the lavaan package in *R*, which is a free and open-source package for estimating a wide range of multivariate statistical models, including all SEM models [11]. In addition, we employed the Unweighted Least Squares and Maximum Likelihood criteria to fit the model, utilizing the BFGS optimization procedure [30,50]. Furthermore, the path diagram of the model in (Equation (1)) is presented in **Figure 3** below[5].

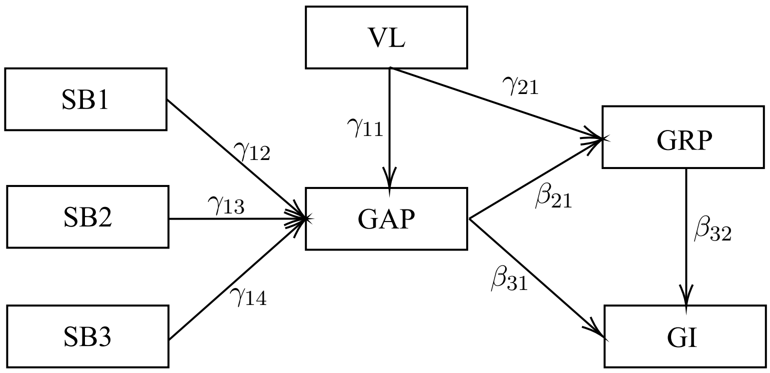

Figure 3. Path diagram of the conceptual model in (Equation (1)).

The next Section 3 will present a detailed analysis of the estimation results.

3. Results and discussion

This section presents the model performance, the model findings, the hypothesis validation, the global discussion, and a comparison of the model findings.

3.1. Model performance

In this part, we aim to validate and assess the predictive performance of our model. To achieve this, the most common various indices used in SEM models are:

1) Comparative Fit Index (CFI): Quantifies the degree to which the hypothesized model fits the observed data [11,16,30,51]. CFI ranges from 0 to 1, with higher values indicating better-fit [51].
2) Tucker Lewis Index (TLI): Assesses the relative fit of the hypothesized model by comparing it with the null model [16,41,52].
3) Goodness-of-Fit-Index (GFI): is used to evaluate how the proposed model fits the observed data by comparing the difference between the empirical covariance matrix and the implied covariance matrix. GFI value higher than 0.95 is accepted [16,51].
4) Adjusted Goodness-of-Fit-Index (AGFI): is a modification of GFI by considering the degrees of freedom in the model, penalizing for model complexity [16,41,52, 53].
5) Root Means Square Error of Approximation (RMSEA): is a fit index that provides how the model fits the observed data per degree of freedom, and it considers the

model complexity. Overall, an RMSEA less than 0.10 implies a poor fit, and an RMSE higher than 0.94 implies a good fit [16,41,53].

6) Standardized Root Mean Square Residual (SRMR): is the Standardized Root Mean Square Residual, a measure of the mean absolute correlation residual, with smaller values revealing a good model fit [11,16,41]. An SRMR value closer to 0 indicates a better fit.

7) Bentler-Bonett Normed Fit Index (NFI): is an incremental fit index that compares the fit of the model to the baseline model. An NFi value closer to 1 indicates a better fit [11,16,41,53].

Table 7 presents the values of these indices:

Table 7. The values of fit indices.

Index	CFI	GFI	AGFI	GFI	TLI	RMSEA	NFI
Value	0.974	0.959	0.836	0.945	0.170	0.018	0.974

Table 7 reveals key insights into the adequacy of the proposed model's fit to the data. With a CFI of 0.974, the model demonstrates a strong comparative fit, indicating that it closely aligns with the observed covariance matrix. Similarly, a GFI at 0.959 suggests a satisfactory overall fit and a higher value of AGFI at 0.836. In addition, TLI at 0.945 means the model fits the data very well, and the RMSEA of 0.170 is acceptable, although it is a little higher than the global threshold used in SEM models, which is 0.11 [41]. Furthermore, an SRMR of 0.018 is extremely low, indicating that the differences between the observed and predicted covariances are very small and that the model captures the relationships between variables very well. Besides, an NFI of 0.974, meaning that the model fits the data very well, explains a significantly greater proportion of the variance in the data than the null model, and that the relationships specified in the model are justified by the data. Hence, the proposed model is statistically significant, highly predictive, and more adequate, in efficiency and reliability in predicting household energy consumption.

After proving the validity and the coherence of our model, we proceed to compare numerically the performance and efficiency between our model and the other models in the previous studies presented in Part 2.2. We show the performance of our model with that of MLR that is purely independent, and which does not take into account indirect effects between variables. Therefore, we compare the values of R-squared (R^2) for the fit that was found for each of the applications of the MLR using lm function in R for each dependent variable separately (Global Active Power, Global Reactive Power, and Global Intensity) with their values found in our model in (Equation (1)), and conducted using lavaan package.

Table 8 below presents these values of R^2.

Table 8. Comparison between PAM, and MLR using R^2.

Dependent variable	PAM R^2	MLR R^2
Global Active Power	0.704	0.704
Global Reactive Power	0.065	0.065
Global Intensity	0.998	0.998

Table 8 shows the R^2 values are equal for both approaches, which is expected since, in simple cases where there is a single dependent variable and several independent variables, PAM and MLR give the same fitting results. As mentioned in

Part 2.2, PAM is an extension of MLR and retains the same fitting performance for an isolated equation. However, its advantage lies in its ability to model indirect effects, taking into account intermediate relationships between variables. For example, it can measure the indirect impact of Sub Metring 1 on Global Intensity via Global Active Power and Global Reactive Power, which is not possible with independent regression models.

After showcasing the performance and efficiency of our model, we proceed to interpret and examine the estimated coefficients. The next part 3.2 is dedicated to this task.

3.2. Model findings and hypotheses validation

In this part, all direct and indirect effects between the variables will be analyzed and examined. In addition, all hypotheses proposed above will be discussed. **Table 9** below presents the coefficients and their *p*-values indicating all variables are significant.

Table 9. Estimates and significance levels.

Variable	Coefficient	*P*-value
Global Active Power		
Voltage	−0.116	0.000
Sub Metering 1	0.370	0.000
Sub Metering 2	0.359	0.000
Sub Metering 3	0.510	0.000
Global Reactive Power		
Global Active Power	0.246	0.000
Voltage	−0.022	0.000
Global Intensity		
Global Active Power	0.994	0.000
Global reactive Power	0.018	0.000

The results in **Table 9** show the direct, and indirect effects of several variables on household energy consumption in the context of the hypothesis proposed. By analyzing these findings, we can reveal how different factors influence the energy parameters in the system.

For hypothesis $\mathbf{H}_1$, the findings confirmed a large part of this hypothesis, with significant coefficients for all variables. However, voltage has a negative effect (-0.116) on the global active power. This negative direct effect shows that if voltage increases, active power tends to decrease slightly, this can be explained by the adjustments in the system to maintain energy balance. This inverse effect between voltage and direct energy consumption is a frequent adjustment in systems where one tries to maintain energy stability with changes in voltage. On the other hand, the sub-metered consumption metrics (Sub Metering 1, Sub Metering 2, and Sub Metering 3) have significant and highly positive effects respectively (0.370, 0.359, and 0.510) on the global active power. This direct impact of these sub-measures on active energy consumption, suggests that each sub-measure adds to the energy consumption in a proportional way to their coefficient, a key point to understanding the individual contributions of each subsector to overall energy consumption.

Concerning $\mathbf{H}_2$, the results indicate that global active power has a significant positive effect (0.246) on global reactive power. This effect supports the idea that the expansion of active power supply leads to an increase in reactive power demand, supporting the proposal of a direct interaction between active and reactive power requirements within the system. In addition, the low negative effect (-0.022) of voltage on global reactive power, suggests a decrease in the reactive power demand when the voltage increases. This inverse effect is said to be the automatic corrections in the system to adjust the voltage fluctuations.

For $\mathbf{H}_3$, the results show that the effect of active global power on global intensity is very high, with a coefficient value very close to (0.994). This impact shows that energy intensity is largely based on active consumption and that it is likely that energy intensity is almost based on the turnaround of active consumption in this system. In addition, global reactive power also has a positive impact on global intensity, which is (0.018), albeit on a smaller scale. In any case, a small contribution from reactive power is contained within the contributions to overall energy intensity, which justifies that variations in reactive power contribute even marginally to the evolution of energy intensity.

For $\mathbf{H}_4$, and $\mathbf{H}_5$, the results can be examined and viewed in a coherent logic. As regards the direct effect of the sub-measures on global active power, it is important to note that it plays an indirect but essential role in meeting reactive power demand and energy intensity. Because of the primacy of active power and therefore of active demand, these sub-measures lack direct power over reactive power as well as energy intensity due to their cumulative effect on active consumption, which clarifies the type of mediated interaction hypothetically expected.

Table 10 below explicitly summarizes the outcomes of the five hypotheses.

Table 10. Summarization of the outcomes for the hypotheses.

Hypothesis	**Outcome**
$\mathbf{H}_1$	
Voltage → (Positively) Global Active Power	Not Confirmed
Sub Metering 1 → Global Active Power	Confirmed
Sub Metering 2 → Global Active Power	Confirmed
Sub Metering 3 → Global Active Power	Confirmed
$\mathbf{H}_2$	
Voltage → (Positively) Global Reactive Power	Not Confirmed
Global Active Power → Global Reactive Power	Confirmed
$\mathbf{H}_3$	
Global Active Power → Global Intensity	Confirmed
Global Reactive Power → Global Intensity	Confirmed
$\mathbf{H}_4$	
Sub Metering (1, 2, and 3) → (Indirectly) Global Reactive Power	Confirmed
$\mathbf{H}_5$	
Sub Metering (1, 2, and 3) → (Indirectly) Global Intensity	Confirmed

Furthermore, **Table 11** below presents the values of indirect effects.

Table 11. Indirect effects values.

Indirect effect	Value
Between Sub Metering (1, 2, and 3), and Global Reactive Power	
Sub Metering 1 $\rightarrow$ (Indirectly) Global Reactive Power	$0.37 \times 0.25 = 0.092$
Sub Metering 2 $\rightarrow$ (Indirectly) Global Reactive Power	$0.36 \times 0.2 = 0.09$
Sub Metering 3 $\rightarrow$ (Indirectly) Global Reactive Power	$0.51 \times 0.25 = 0.13$
Between Sub Metering (1, 2, and 3), and Global Intensity	
Sub Metering 1 $\rightarrow$ (Indirectly) Global Intensity	$0.37 \times (0.99 + 0.25 \times 0.018) = 0.37$
Sub Metering 2 $\rightarrow$ (Indirectly) Global Intensity	$0.36 \times (0.99 + 0.25 \times 0.018) = 0.358$
Sub Metering 3 $\rightarrow$ (Indirectly) Global Intensity	$0.51 \times (0.99 + 0.25 \times 0.018) = 0.50$

Finally, **Figure 4** below visualizes the path diagram of the model with estimated coefficients and calculated indirect effects.

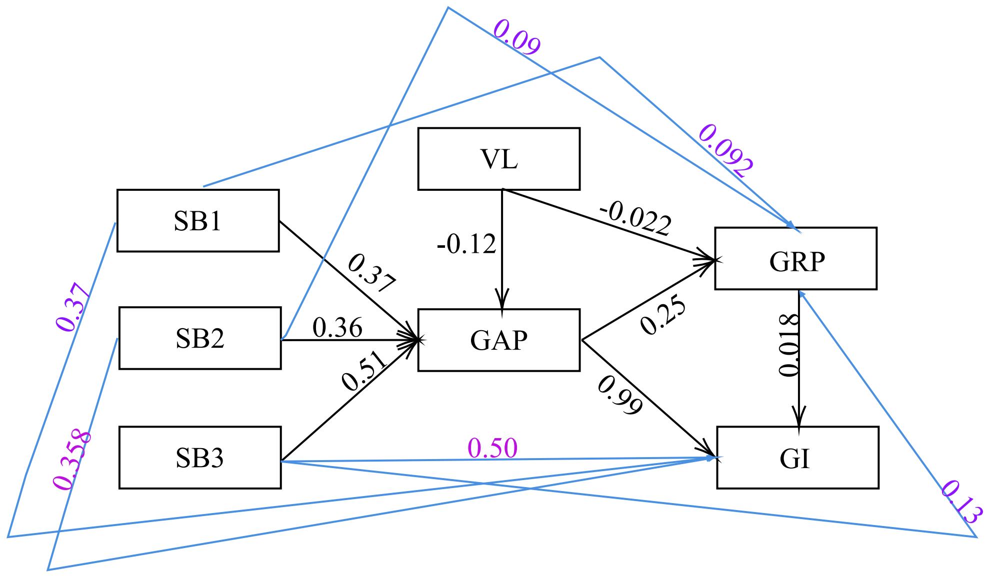

Figure 4. Path diagram of the model with estimated coefficients and calculated indirect effects.

The next Part 3.3 discusses these findings and compares them with the findings of previous studies.

3.3. Discussion

Our model findings provide insightful revelations about household electricity consumption and offer a detailed view of how energy consumption behaviors are interconnected in a residential context. Our model has made it possible to break down direct and indirect effects, making it an appropriate tool for research into domestic energy consumption. Compared with MLR, our model takes account of complex interactions and chain effects. This confirms the findings of [10, 11, 19, 21, 54] that more advanced analytical tools are needed to understand energy consumption in high-intensity urban areas and as climate change progresses. These advanced analytical tools also provide a basis for making informed policy decisions by identifying the most significant impacts on energy behavior in the home and developing specific measures to address them [10, 11]. Each examined variable in our model illuminates various aspects of energy consumption offering a robust foundation for crafting targeted policy interventions.

Sub Metering components (1, 2, et 3) significantly influence global reactive power demand indirectly via their effects on global active power. These findings align with the findings of [10,11] which underline the need for better management of consumption sub-segments to relieve pressure on limited sources of available energy. By targeting sub-measures, energy efficiency policies can become more precise and provide more in-depth knowledge, making it possible to optimize the use of domestic energy and reduce reactive energy [9–11,19–21].

In addition, the positive effect of the sub-measures (Sub Metering 1, 2, and 3) such as cooking or space heating/cooling or heating water on the global power (active and reactive) suggests that each undervalued source contributes proportionately to global energy consumption. This outcome aligns with the findings of [9, 22] that identified portable electric heating and electric water heating as significant drivers of high electricity demand in UK homes, and also the studies of [2] that said that space heating/cooling, domestic hot water, and cooking represent, respectively 30%, 12%, and 7% of global consumption in Tours (France). Consequently, appropriate bespoke energy interventions could be made to optimize the use of critical sub-segments, which may be more effective in high-cost random resource environments [10,11,19].

As shown in Figure 4, each Sub Metering (1, 2, et 3) also has a significant indirect effect with different scales on the global intensity, which means that each household's ownership of electrical appliances will indirectly affect electricity consumption. This aligns with the outcome of [55] that announces that the household's ownership of electrical appliances alone will not affect electricity consumption, but the greater the number of appliances owned, the more opportunities that exist for electricity use in UK homes.

Furthermore, the voltage has a significant negative impact on global power (active and reactive). While these findings are consistent with studies of [56] that present the effects of voltage reductions on real power, reactive power, and energy for individual and composite loads, other studies could explore whether this relationship holds across diverse climatic or socio-economic conditions [55]. This negative effect of voltage and global power can be understood through physical and behavioral mechanisms. Also, the fact that higher voltages can ensure efficient operation of appliances, consequently reducing energy demand [10,11,19].

Finally, **Figure 4** demonstrates that the main components (global reactive power, global active power, voltage, global intensity, Sub Metering 1, Sub metering 2, and Sub-Metering 3) have a considerable impact directly and indirectly on electricity demand, confirming and reinforcing the previous studies [10,11,19,22].

The next Section 4 presents the conclusions, limitations of the study and discusses directions for future research.

4. Conclusions

This study addresses the challenge of reducing household power consumption, a critical task for achieving efficient and sustainable energy management. Reducing household energy use is essential for conserving financial resources and protecting the environment. Our approach involved examining household power consumption patterns and identifying unusual behaviors to achieve SDG 7's objectives. We have proposed several hypotheses to model household power consumption usage and behaviors based on the analysis of a dataset of over 2 million users spanning four years, by considering correlations between variables and existing studies on energy consumption and power demand management. In addition, we have discussed in depth the models used to study these hypotheses, highlighting their limitations in capturing causal and mediating relationships. After, we have proposed a PAM that overcomes these limitations by capturing and understanding the components of the direct and indirect effects on household energy consumption. Numerical and theoretical comparisons that demonstrated the advantages and efficiency of PAM are also given. The results show that Sub-metering components associated with specific uses, like cooking or water heating, have significant indirect impacts on global intensity through active power and that the voltage has a negative effect on the global power (active and reactive) due to the physical and behavioral mechanisms. Our findings provide an in-depth understanding of household electricity power consumption. This will improve forecasting and enable real-time energy management tools, extending to the design of precise energy efficiency policies.

This study presents certain limitations, providing perspectives for future research. Firstly, the scope of the data is residence *Sceaux* (7 km of Paris, France), which may not fully cover France's household electric power consumption challenges and opportunities. Secondly, the temporal aspect of the dataset, with the data being 14 years old. To overcome these limitations, two works are in progress: (1) Extending research to a national scale can yield insights into regional variances and enable a comprehensive strategy for the country; (2) An exhaustive approach that updates this dataset with a complement of socioeconomic variables to ensure its relevance and accuracy and combines PAM with advanced time series models [57] will offer a deep, comprehensive analysis of household electric power consumption trends, a life-cycle theory analysis of French household electricity demand, and the determinants of high electrical energy demand.

In addition, a current work integrates PAM with confirmatory factor analysis [51] to identify the socio-economic, environmental, and dwelling factors contributing to high electrical energy demand in Moroccan domestic buildings.

Moreover, another current work that combines Recurrent Neural Networks (RNNs), Multiple Correspondence Analysis [7], Ascending Hierarchical Classification with SEM models to identify distinct categories of households based on the electrical energy demand in Mauritania domestic buildings by analyzing complex direct and indirect effects between observed socio-economic variables such as household income and education level and latent variables such as environmental awareness. This approach will enhance our understanding of household electric power consumption challenges in African cities.

Finally, an application for the detection and localization of household energy distribution system leaks in real-time in Ben Guerir city (Morocco) [58] using wavelet decomposition, machine learning, and SEM models is in progress. Also, this application that contains smart grid applications to flatten spikes in electricity will reduce the overall electricity consumption by applying Demand-Side Management (DSM) strategies [2].

Author contributions: Conceptualization, SAEA, MB, JC, EBD, RA, MA and MH; methodology, SAEA; software, SAEA; validation, SAEA, EBD, JC, RA, MA, IS and MH; formal analysis, SAEA, EBD, IS and MH; data curation, SAEA and MH; writing—original draft preparation, SAEA, MB and IS; writing—review and editing, SAEA, EBD, MB, RA, MA and IS; supervision, JC; project administration, MB and JC. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: The authors gratefully acknowledge the support of editors during the publication process as well as the useful feedback provided by two anonymous referees.

Conflict of interest: The authors declare no conflict of interest.

Notes

[1] A direct effect is a causal relationship between two variables that is not mediated by any other variable. It is represented by a unidirectional arrow and quantified by the coefficient connecting these variables [30].

[2] The indirect effect between two variables is an effect mediated by one or more variables and measured as the product of the coefficients from the first one to the last one [30].

[3] Variable that is not influenced by other variables in the model [39].

[4] Variable that is influenced by one or more variables in the model [39].

[5] Where (SB1, SB2, and SB3) respectively represent Sub Metering 1, Sub Metering 2, and Sub Metering 3, and (VL, GAP, GRP, and GI) represent respectively Voltage, Global Active Power, Global Reactive Power, and Global Intensity.

References

1. Elavarasan. R, et al. (2023). The untold subtlety of energy consumption and its influence on policy drive towards Sustainable Development Goal 7. Applied Energy.
2. Falaki. F, et al. (2021). A Spatio-Temporal Analysis of electric appliance end-use demand in the residential sector: Case study of Tours (France). Sustainable Cities and Society.
3. IEA Report (2024).
4. Belaïd. F, et al. (2022). A life-cycle theory analysis of French household electricity demand. Journal of Evolutionary Economics.
5. Popescu. C, et al. (2022). Energy transition in European Union—Challenges and opportunities. Energy Transition: Economic, Social and Environmental Dimensions.
6. Valavanidis. A. (2024). Global Electricity Generation from Renewable Sources.
7. Lévy. J, et al. (2018). The determinants of domestic energy consumption in France: Energy modes, habitat, households and life cycles. Renewable and Sustainable Energy Reviews.
8. Weber. C, et al. (2000). Modeling lifestyle effects on energy demand and related emissions. Energy policy.
9. Belaïd. F, et al. (2016). Understanding the spectrum of domestic energy consumption: Empirical evidence from France. Energy Policy.
10. Krishnan. L, et al. (2023). Prediction of electrical power consumption in the household: fresh evidence from machine learning approach. Energy Efficiency.
11. Tran. L, et al. (2022). Influence of household factors on energy use in Vietnam based on path analysis. Journal of Building Engineering.
12. Valenzuela. C, et al. (2014). An analysis of monthly household energy consumption among single-family residences in Texas. Energy Policy.
13. Hebrail. G, et al. (2006). UCI Machine Learning Repository. https://doi.org/10.24432/C58K54.
14. Frank. A, et al. (2010). Uci machine learning repository. Irvine, ca: University of California, School of Information and Computer Science. http://archive. ics. uci. edu/ml.
15. Hebrail. G, et al. (2012). Individual household electric power consumption Data Set (2012). `http://bit.ly/1byLbr0`.
16. Verma. P, et al. (2023). Analysing Path Analysis with Multiple Regression. In Understanding Structural Equation Modeling: A Manual for Researchers (pp. 91-113). Cham: Springer International Publishing.
17. Best. D, et al. (1975) Algorithm AS 89: the upper tail probabilities of Spearman's rho." Journal of the Royal Statistical Society. Series C (Applied Statistics)
18. Pearson. S. (1931). The test of significance for the correlation coefficient. Journal of the American Statistical

Association.
19. Huebner. G, et al. (2016). Understanding electricity consumption: A comparative contribution of building factors, socio-demographics, appliances, behaviors, and attitudes. Applied Energy.
20. Huebner. G, et al. (2013). What role do comfort, habit, and knowledge about the heating system play?. Energy and buildings
21. Huebner. G, et al. (2015) Explaining domestic energy consumption–the comparative contribution of building factors, socio-demographics, behaviors, and attitudes. Applied Energy.
22. Jones. R, et al. (2016). Determinants of high electrical energy demand in UK homes: Appliance ownership and use. Energy and Buildings.
23. Ahn. J, et al. (2002). Beyond single equation regression analysis: Path analysis and multi-stage regression analysis. American journal of pharmaceutical education.
24. Mishra. D, et al. (2010). Analyzing the relationship between dependent and independent variables in marketing: A comparison of multiple regression with path analysis. Available at SSRN.
25. Garson. G. (2013). Path analysis. Asheboro, NC: Statistical Associates Publishing.
26. Cohen. J, et al. (2013). Applied multiple regression/correlation analysis for the behavioral sciences. Routledge.
27. Mueller. R. (1996). Linear Regression and Classical Path Analysis. In: Basic Principles of Structural Equation Modeling. Springer Texts in Statistics. Springer.
28. Streiner. D. (2005). Finding our way: an introduction to path analysis. The Canadian Journal of Psychiatry.
29. Turner, et al (1959). The regression analysis of causal paths. Biometrics.
30. Ebnou. Seyid, Iaousse. M, El Hadri. Z. (2023) New lights on the correlation matrix implied by a recursive path model. Qual Quant.
31. Wang. Y. (2005). A multinomial logistic regression modeling approach for anomaly intrusion detection. Computers & Security.
32. Wang. F, et al. (2018). Residential electricity consumption level impact factor analysis based on wrapper feature selection and multinomial logistic regression. Energies.
33. Kayacı. M. (2023). Ensemble Machine Learning Approaches for Prediction of Türkiye's Energy Demand. Energies.
34. Shahzad. A, et al. (2024). New ridge parameter estimators for the quasi-Poisson ridge regression model. Sci Rep.
35. Hilt. D, et al. (1977). Ridge, a computer program for calculating ridge regression estimates. Department of Agriculture, Forest Service. Northeastern Forest Experiment Station.
36. Feng. X, et al. (2013). Study on China's Energy Demand Ridge Regression Prediction Based on Path Analysis. Advanced Materials Research.
37. Yu. F, et al. (2024). Predicting and decarbonizing carbon emissions from building energy use in Hong Kong: A LASSO regression approach. Energy for Sustainable Development.
38. Bouznit. M, et al. (2023). Economic growth, human capital, and energy consumption in Algeria: evidence from cointegrating polynomial regression and a simultaneous equations model. Environ Sci Pollut Res.
39. Ebnou. Seyid, Iaousse. M, El Hadri. Z. (2022). Generalizing the Properties of the Finite Iterative Method for the Computation of the Covariance Matrix Implied by a Recursive Path Model. Statistics, optimization, and information computing.
40. Bollen. A. (1989). Structural equations with LV. Wiley-Interscience.
41. Kline. R. (2016). Principles and practice of structural equation modeling. The Guilford Press, 2016.
42. Wright. S. (1921). Correlation and causation, Journal of Agricultural Research.
43. Sahli. A, et al. (2024). Extending El-Hadri-Sahli-Hanafi procedure for path analysis with non-standardized variables. Qual Quant.
44. Hong. J. et al. (2016). A multi-regional structural path analysis of the energy supply chain in China's construction industry. Energy Policy.
45. Gui. S, et al. (2017). Path analysis of factors impacting China's CO2 emission intensity: viewpoint on energy. Energy policy.
46. Irfan. M, et al. (2021). Consumers' intention-based influence factors of renewable energy adoption in Pakistan: a structural equation modeling approach. Environmental Science and Pollution Research.
47. Elahi. E, et al. (2022). Understanding farmers' intention and willingness to install renewable energy technology: A solution to reduce the environmental emissions of agriculture. Applied Energy.
48. Chen. W, et al. (2017). Analysis of the impact path on factors of China's energy-related CO2 emissions: a path analysis with latent variables. Environ Sci Pollut Res.

49. Babar. Z, et al. (2022). Consumer's perception towards electricity theft: A case study of Islamabad and Rawalpindi using a path analysis. Energy policy.
50. El Hadri. Z, et al (2022). Simple and fast convergent procedure to estimate recursive path analysis model. Behaviormetrika.
51. Seyid. E, et al (2023). Exploring Urban Livability Factors: A Novel Algorithm for Improved Confirmatory Factor Analysis. In 2023 14th International Conference on Intelligent Systems: Theories and Applications (SITA). IEEE.
52. Sathyanarayana. S, et al. (2024). Fit indices in structural equation modeling and confirmatory factor analysis: Reporting guidelines. Asian Journal of Economics, Business and Accounting.
53. Gunzler. D, et al. (2021). Structural equation modeling for health and medicine. Chapman and Hall/CRC.
54. Rojek. I, et al. (2024). Green Energy Management in Manufacturing Based on Demand Prediction by Artificial Intelligence—A Review. Electronics.
55. Lomas. R, et al. (2015). Determinants of high electrical energy demand in the UK. Database.
56. Scalley. D, et al (1981). The Effects of Distribution Voltage Reduction on Power and Energy Consumption. IEEE Transactions on Education.
57. Kim. H, et al. (2023). Time-series clustering and forecasting household electricity demand using smart meter data. Energy Reports.
58. Adraoui. M, et al. (2024). A two-phase approach for leak detection and localization in water distribution systems using wavelet decomposition and machine learning. Computers & Industrial Engineering.