Income-Based Energy Transition in Chilean Households:

Extending Fercovic et al. (2024) with 2022 Chilean Household Data and an Emphasis on Rural-Urban Heterogeneity

Isaac Qi and Yunke Li

FRE 528, MFRE, University of British Columbia

December 5, 2025

Abstract

This paper extends Fercovic et al. (2024) by examining whether the identified income-firewood relationship from 2015-2017 in Chile still holds true in 2022. The original paper found an economically trivial income elasticity of firewood use (about 0.1 percentage points lower probability of using firewood for a 10% income increase), implying that passive income growth is unlikely to drive a meaningful energy transition. Using CASEN household survey data for 2015, 2017, and 2022 combined with commune-level forest and climate variables from Hansen et al. (2013), we estimate pooled probit models with year dummies, income-year interactions, and rural heterogeneity. We continue to focus on heating households and treat CASEN as repeated cross-sections rather than a true panel. Across all specifications, the income effect remains statistically significant but economically trivial, and this pattern is stable through 2022. Our results suggest that even in a high-income and fast-growing economy, income growth alone is not enough to move households away from firewood and especially in rural areas, therefore energy transition policies remain necessary.

1 Introduction

Fercovic et al. (2024) examine whether income growth helps Chile move away from firewood as a primary heating fuel. Using CASEN 2015 and 2017 survey data supplemented with climate and forest variables, they find that a 10% income increase lowers the probability of using firewood by only about 0.1 percentage points (one-tenth of a percentage point). The relationship is statistically significant but economically trivial, so simply getting richer does not solve Chile's air quality and health consequences linked to firewood use. This finding runs against the simple energy ladder framework where higher incomes naturally push households from biomass to cleaner fuels. The original paper already includes well selected controls and robustness checks, and it concludes that policy needs to go beyond relying on income growth alone. What it cannot do, however, is say whether this weak income effect is a one-off feature of 2015-2017 or a more stable pattern.

Our project fills that gap by adding the latest CASEN wave (2022)and asking a simple question: has the income-firewood relationship changed by 2022? We focus on two motivations. First, Chile's GDP per capita rose slightly from 14879.9 (USD) to 15405.6 (USD) between 2017 and 2022, and while volatility is observed due to the COVID-19 pandemic, the overall trend is positive (World Bank, 2025). Second, 2022 captures post-COVID recovery, a period when budgets and energy needs were disrupted. This provides an window into how Chile responds to economic shocks in terms of fuel choice, though we acknowledge that this would be speculative and exploratory.

We contribute in two main ways. Empirically, we provide a simple but important temporal validation of the original result by pooling 2015, 2017, and 2022 data and comparing income effects across waves. Methodologically, we show how to merge multi-wave survey data with commune-level forest and climate variables. Looking ahead to policy, we ask whether five more years of growth have made income a more powerful driver of fuel switching, or whether the original pessimistic conclusion still holds.

2 Extension: Temporal Validation and Rural-Urban Heterogeneity

Our project introduces two main extensions to Fercovic et al. (2024). First, we add the 2022 CASEN household survey and pool it with the 2015 and 2017 waves. This five-year extension lets us test whether the income-firewood elasticity changed as Chile's economy grew further and as households adjusted after COVID-19. In practice, we baseline our probit model with the 2015-2017 dataset, and incorporate 2022 to assess whether any changes appear in either coefficients or predicted probabilities. Second, we explicitly compare rural and urban households by estimating separate probit models for each subsample and a pooled model with Rural $\times$ Year interactions. This zoom-in allows us to ask whether income growth and forest availability matter differently in rural versus urban areas, and whether the large rural-urban firewood gap has narrowed or persisted by 2022.

Theoretical motivation comes from the energy ladder hypothesis, which predicts that as income rises, households progressively adopt cleaner fuels. Temporal validation matters because income elasticities may be non-linear, and Chile's 2017-2022 income growth could push more households past switching thresholds, strengthening the observed elasticity even if it was weak in 2015-2017. In addition, it takes time for households at an income level that can afford the energy transition to make the switch, and if there are non-linearity in the rate of adoption, the 2022 data has potential of capturing such changes despite not being a true panel. Energy choices also exhibit inertia, so a weak elasticity in one period might strengthen as infrastructure for alternatives such as natural gas networks and electric heat pumps expands.

Policy relevance is straightforward. If the elasticity strengthened by 2022, passive income growth policies might eventually work, just more slowly than hoped. If it remained constant or weakened, Chile must rely on targeted subsidies, regulations, and infrastructure rather than waiting for income growth to do the work. The rural-urban zoom-in is crucial for targeting: if rural households remain heavily reliant on firewood at all income levels, then infrastructure and supply-side interventions become necessary.

Fercovic et al. (2024) use forest cover and loss from Hansen et al. (2013), processed from 30m raster data. Extending these rasters through 2020-2022 would require substantial computing time and storage, which is beyond the scope of this project (we've attempted, but failed after 5 hours of compiling). We will further discuss these variables in the data section, and return to this limitation in the robustness and limitations sections.

3 Data

3.1 Primary Data: CASEN Household Surveys

We use Chile's Encuesta de Caracterización Socioeconómica Nacional (CASEN) for 2015, 2017, and 2022 waves. CASEN is a nationally representative cross-sectional survey covering socio-demographics, income, housing, and energy use. Our analysis focuses on households with heating systems, following Fercovic et al. (2024)'s selection framework.

3.2 Forest Covariates: Hansen Global Forest Change

We use forest cover and loss data from Hansen et al. (2013) included with the dataset provided by Fercovic et al. (2024), which provides 30m-resolution satellite imagery tracking global deforestation. We obtain processed forest data from Fercovic et al. (2024)'s replication package, aggregated to the commune level. The forest measures include forest cover in km^2 and forest loss in km^2.

For 2022 analysis, we treat 2019 forest cover and cumulative loss as time-invariant proxies. While this introduces measurement error, we found that forest cover is highly persistent at the commune level with correlation exceeding 0.99 between 2015 and 2017, therefore 2019 should be a good proxy for 2022. For forest loss, we found that there were moderate within-commune changes between 2015 and 2017 with a correlation of 0.67, but ultimately using 2019 data is still an improvement compared to shifting 2017's forest data to 2022.

3.3 Temperature Covariates

Minimum temperature is highly persistent at the commune level with correlation exceeding 0.99 between 2015 and 2017, therefore we took the mean and carried it forward to 2022.

3.4 Key Variables

The dependent variable `Firewood` is a binary indicator equal to 1 if firewood is the primary heating fuel and 0 otherwise. The main independent variable is `Log_Income`, the log of total household income. We include polynomial terms $\texttt{Log\_Income}^2$ and $\texttt{Log\_Income}^3$ to capture non-linear income effects. Climate is captured by `Min_Temp`, the mean minimum winter temperature in degrees Celsius. Household size is measured by `n_Household`, the number of household members. Housing tenure is represented through binary indicators `Owner`, `Renter`, `Ceded`, and `Other`, where `Owner` equals 1 if the household owns the dwelling, `Renter` equals 1 if the household rents, `Ceded` equals 1 if the dwelling is ceded or granted, and `Other` equals 1 for all other ownership types. Location is captured by `Rural`, a binary indicator equal to 1 if the household is in a rural area. Forest variables include `Forest`, the total forest cover in the commune measured in km^2, and `Forest_Loss`, the cumulative forest loss in the commune since 2001 also measured in $k\ m^2$. Finally, `Year` indicates the survey year with values of 2015, 2017, or 2022.

3.5 Descriptive Statistics

The summary statistics table (Table 1) presents descriptive statistics by survey year for the trimmed analysis sample. Mean household income surveyed rises substantially between 2015 and 2022, from 817,985 to 1,097,713 Chilean pesos, consistent with Chile's continued economic development. The standard deviation of income remains high, ranging from 612,807 pesos in 2015 to 665,249 pesos in 2022. Minimum temperatures and household size are stable across survey waves, and forest variables change only slightly on average. The shallow change in firewood use relative to income growth foreshadows our later finding of a stable, economically trivial income elasticity.

Table 1: Descriptive Statistics by Year

Year	Variable	N	Mean	SD	Min	Max
2015	Income	73,450	817,985	612,807	0	20,555,833
2017	Income	62,354	887,392	619,334	0	13,950,000
2022	Income	64,151	1,097,713	665,249	0	13,742,084
2015	Log_Income	73,450	13	1	0	17
2017	Log_Income	62,354	13	1	0	16
2022	Log_Income	64,151	14	1	0	16
2015	Min_Temp	73,450	7	2	-4	12
2017	Min_Temp	62,354	7	2	-4	12
2022	Min_Temp	64,151	7	2	-4	12
2015	n_Household	73,450	3	2	1	15
2017	n_Household	62,354	3	2	1	16
2022	n_Household	64,151	3	1	1	12
2015	Owner	73,450	1	0	0	1
2017	Owner	62,354	1	0	0	1
2022	Owner	64,151	1	0	0	1
2015	Renter	73,450	0	0	0	1
2017	Renter	62,354	0	0	0	1
2022	Renter	64,151	0	0	0	1
2015	Ceded	73,450	0	0	0	1
2017	Ceded	62,354	0	0	0	1
2022	Ceded	64,151	0	0	0	1
2015	Other_Ownership	73,450	0	0	0	1
2017	Other_Ownership	62,354	0	0	0	1
2022	Other_Ownership	64,151	0	0	0	1
2015	Rural	73,450	0	0	0	1
2017	Rural	62,354	0	0	0	1
2022	Rural	64,151	0	0	0	1
2015	Forest	73,450	39,294	163,323	0	2,122,514
2017	Forest	62,354	52,398	204,245	0	2,122,514
2022	Forest	64,151	44,177	176,042	0	2,122,514
2015	Forest_Loss	73,450	213	389	0	2,402
2017	Forest_Loss	62,354	197	359	0	2,402
2022	Forest_Loss	64,151	202	361	0	2,402

Figure 1 visualizes firewood use by income quintile across the three survey waves. This figure is constructed by dividing households with heating systems into five equal-sized income groups and calculating the share using firewood as primary heating fuel within each quintile for each year. The bars show raw proportions without controlling for other covariates. The negative income gradient is evident in all years, with higher-income households consistently using less firewood. However, the gradient did not steepen noticeably between 2015-2017 and 2022. The near-parallel pattern across survey years provides visual evidence that income elasticity remained stable despite substantial income growth.

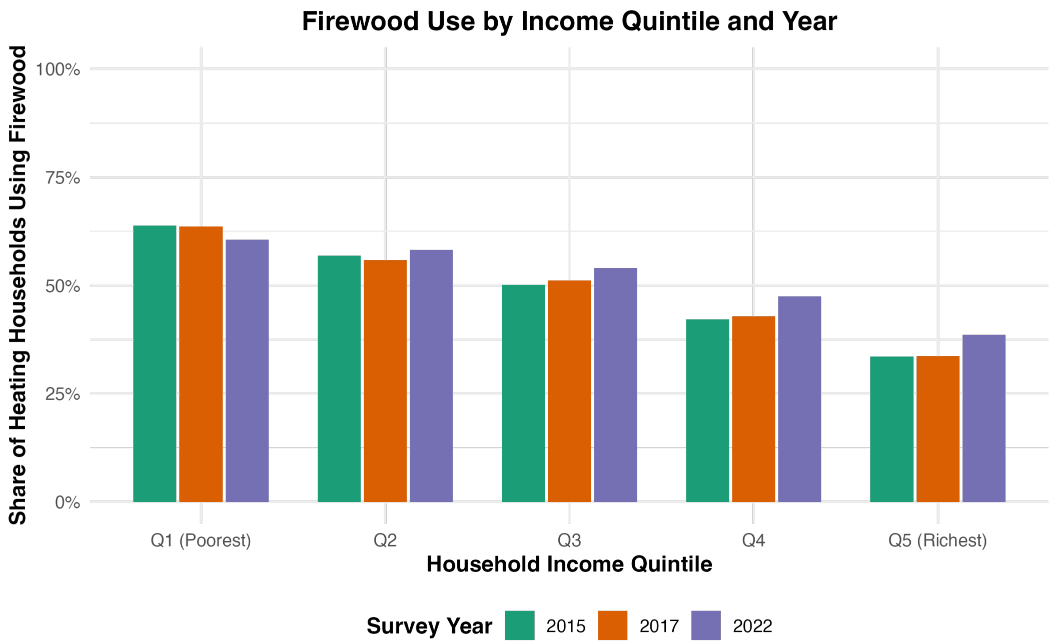

Figure 1: Firewood Use by Income Quintile and Year (2015, 2017, 2022)

3.6 Data Cleaning and Sample Selection

Following Fercovic et al. (2024)'s approach, we construct a trimmed analysis sample in several steps. We start from household-level CASEN data for 2015, 2017, and 2022 with temperature and forest variables. We then compute log income and restrict the main probit analysis to households that chose to heat, keeping non-heating households only for descriptive comparisons. We trim extreme values of log income and minimum temperature at the 5th and 95th percentiles and household size at the 1st and 99th percentiles. This is consistent across all versions of our analysis. Finally, we drop communes with missing temperature data after the merge, the result of additional communes surveyed in 2022 that were not included in 2015 or 2017.

The trimmed dataset contains 73,450 observations in 2015, 62,354 in 2017, and 64,151 in 2022, for a total of 199,955 household-year observations as reported in the summary statistics table. Consistent with Fercovic et al. (2024), we find that CASEN is a repeated cross-section rather than a true panel, as household IDs do not repeat across survey years. Consequently, we employ pooled cross-sectional models with year fixed effects rather than panel methods.

4 Methods

4.1 Baseline Model: Probit on Heating Households

Our baseline specification follows closely with the structure of Fercovic et al. (2024)'s model barring minor differences in variable choice but simplifies the selection step, where we did not replicate the 2-stage Heckman procedure. Instead, we filtered for only households that report using any kind of heating and estimate a single probit for choosing firewood as their primary heating fuel:

$$
\begin{aligned}
\Pr(\text{Firewood}_i = 1 \mid X_i) = \Phi\Big(&\beta_0 + \beta_1 \ln Y_i + \beta_2 (\ln Y_i)^2 + \beta_3 (\ln Y_i)^3 \\
&+ \beta_4 T_i + \beta_5 H_i + \beta_6 \text{Owner}_i + \beta_7 \text{Ceded}_i + \beta_8 \text{Other}_i \\
&+ \beta_9 \text{Rural}_i + \beta_{10} F_i + \beta_{11} L_i + \beta_{12} (\text{Rural}_i \times L_i) \\
&+ \beta_{13} (\ln Y_i \times T_i) + \theta_{2017}\, \text{Year}_{2017,i}\Big).
\end{aligned}
\tag{1}
$$

where $\Phi(\cdot)$ is the standard normal CDF, $\ln Y_i$ is log household income, T_i is minimum temperature, H_i is household size, F_i is forest cover, and L_i is forest loss. $\text{Year}_{2017,i}$ is a dummy variable equal to 1 if household i belongs to that survey year, and 0 otherwise, with 2015 serving as the omitted reference category.

4.2 Extension Model 1: Temporal Change in Income Elasticity

To test whether the income effect changed over time, we augment the baseline with year dummies and pool additional survey waves. The equation below characterizes the model for all three years surveyed years:

$$
\begin{aligned}
\Pr(\text{Firewood}_i = 1 \mid X_i) = \Phi\Big(&\beta_0 + \beta_1 \ln Y_i + \beta_2 (\ln Y_i)^2 + \beta_3 (\ln Y_i)^3 \\
&+ \beta_4 T_i + \beta_5 H_i + \beta_6 \text{Owner}_i + \beta_7 \text{Ceded}_i + \beta_8 \text{Other}_i \\
&+ \beta_9 \text{Rural}_i + \beta_{10} F_i + \beta_{11} L_i + \beta_{12} (\text{Rural}_i \times L_i) \\
&+ \beta_{13} (\ln Y_i \times T_i) + \theta_{2017}\, \text{Year}_{2017,i} + \theta_{2022}\, \text{Year}_{2022,i}\Big).
\end{aligned} \tag{2}
$$

where $\text{Year}_{2017,i}$ and $\text{Year}_{2022,i}$ are dummy variables equal to 1 if household i belongs to that survey year, and 0 otherwise, with 2015 serving as the omitted reference category.

4.3 Extension Model 2: Rural Heterogeneity Over Time

Rural households face distinct energy transition barriers including limited natural gas infrastructure and proximity to forests making firewood more accessible and thus increased reliance. To capture this, we estimate rural and urban subsamples separately and benchmark this to the 2022 extension model. Rural variables are removed from the model as the subsamples are now homogenous.

$$
\begin{aligned}
\Pr(\text{Firewood}_i = 1 \mid X_i) = \Phi\Big(&\beta_0 + \beta_1 \ln Y_i + \beta_2 (\ln Y_i)^2 + \beta_3 (\ln Y_i)^3 \\
&+ \beta_4 T_i + \beta_5 H_i + \beta_6 \text{Owner}_i + \beta_7 \text{Ceded}_i + \beta_8 \text{Other}_i \\
&+ \beta_9 F_i + \beta_{10} L_i + \beta_{11} (\ln Y_i \times T_i) \\
&+ \theta_{2017}\, \text{Year}_{2017,i} + \theta_{2022}\, \text{Year}_{2022,i}\Big).
\end{aligned} \tag{3}
$$

This equation is estimated once for rural households and once for urban households. Differences in coefficients across the two subsamples indicate how the income-firewood relationship and other determinants vary by settlement type and whether these differences change between 2015, 2017, and 2022.

4.4 Estimation Specifications

We use maximum likelihood probit estimation for binary choice of firewood versus other heating methods. CASEN uses rotating households rather than following the same units over time, so we estimate pooled cross-sectional models with year fixed effects instead of panel models. Log income, minimum temperature, and forest loss are mean-centered to minimize multicollinearity. All estimation is conducted in R 4.5.0 using the `glm()` function for probit estimation and the `margins` package for marginal effects.

5 Results

5.1 Baseline Replication (2015 and 2017)

Table 2 presents the comparison between the baseline probit model 1 and the extension model 2. Model 1 contains 105,874 observations from the 2015 and 2017 waves, while Model 2 contains 155,755 observations including 2022 data. Our baseline Model 1 replicates Fercovic et al. (2024)'s core finding: the income coefficients confirm an economically trivial relationship. In Model 1, the linear log-income term is -0.395 with a standard error of 0.016, and in Model 2 it is -0.364 with a standard error of 0.013. These coefficients are similar in magnitude to those reported in Fercovic et al. (2024), who find income elasticities that translate to approximately 0.1 percentage point reduction in firewood probability for a 10% income increase. Higher-order polynomial terms are also highly significant in both models, with quadratic term -0.044 (standard error 0.012) and cubic term 0.081 (standard error 0.018) in Model 1. When translated into marginal effects at the mean, a 10% income increase reduces the probability of using firewood by only a few hundredths of a percentage point, economically tiny despite very precise estimates. The cubic specification captures non-linearities, but these do not overturn the basic conclusion that income growth alone barely moves firewood use, consistent with the original paper's pessimistic assessment of passive income-driven energy transition.

Rural households have a very large positive coefficient on the rural dummy of 1.259 in Model 1 with a standard error of 0.014, and 1.276 in Model 2 with a standard error of 0.012, indicating much higher firewood use probabilities than urban households even after controlling for income, temperature, housing tenure, and forest variables. This rural effect is consistent with Fercovic et al. (2024), who also find a large positive rural coefficient, though our pooled specification with 2022 data shows a slightly larger magnitude. Colder communes are more likely to use firewood, with the minimum-temperature coefficient of -0.111 in Model 1 with standard error of 0.003, and -0.109 in Model 2 with standard error of 0.002, implying that temperature plays a substantial role in firewood dependence. Both forest cover and forest loss are positively associated with firewood use, with the forest loss coefficient of 0.002 and standard error of 0.00002 in both models, consistent with local biomass supply making firewood more attractive. The interaction between log income and minimum temperature is positive with a coefficient of 0.068 in Model 1 and 0.066 in Model 2 (both with standard error 0.004), suggesting that the negative income effect on firewood use is slightly attenuated in colder areas where heating needs are greater, though the effect remains economically modest.

Model 2 pools all three waves with 155,755 observations including 2015, 2017, and 2022 data. This allows us to check whether the baseline relationship between income and firewood use shifted over time. The year dummies show modest temporal shifts: Year 2017 has a coefficient of 0.016 (standard error 0.009), while Year 2022 has a coefficient of 0.071 (standard error 0.009), both relative to the 2015 baseline. The 2022 coefficient is statistically significant and indicates a small increase in firewood use probability relative to 2015 after controlling for covariates, though economically modest. The income coefficients across models remain very similar, with the linear log-income term of -0.364 in Model 2 versus -0.395 in Model 1, so the implied income semi-elasticity remains economically tiny across specifications. This

Table 2: Model Progression from M1 to M2

	Dependent variable:	
	Firewood (1 = uses firewood)	
	M1 (2015–2017)	M2 (2015-2022)
	(1)	(2)
Log Total Household Income	−0.395***	−0.364***
	(0.016)	(0.013)
Log Total Household Income 2	−0.044***	−0.055***
	(0.012)	(0.010)
Log Total Household Income 3	0.081***	0.067***
	(0.018)	(0.015)
Minimum Temperature	−0.111***	−0.109***
	(0.003)	(0.002)
Household Size	0.108***	0.103***
	(0.003)	(0.003)
Owner	0.462***	0.502***
	(0.013)	(0.011)
Ceded	0.334***	0.386***
	(0.016)	(0.013)
Other Ownership	0.420***	0.476***
	(0.031)	(0.023)
Rural	1.259***	1.276***
	(0.014)	(0.012)
Forest Cover	0.00000***	0.00000***
	(0.00000)	(0.00000)
Forest Loss	0.002***	0.002***
	(0.00002)	(0.00002)
Year = 2017	0.018**	0.016*
	(0.009)	(0.009)
Year = 2022		0.071***
		(0.009)
Rural × Forest Loss	−0.001***	−0.001***
	(0.00004)	(0.00003)
Log Income × Min Temperature	0.068***	0.066***
	(0.004)	(0.004)
Constant	−1.342***	−1.369***
	(0.018)	(0.015)
Observations	105,874	155,755
Log Likelihood	−52,862.720	−77,408.360
Akaike Inf. Crit.	105,755.400	154,848.700
McFadden R^2	0.279	0.283
Note:		*p<0.1; **p<0.05; ***p<0.01

stability in the income effect across 2015-2022 validates Fercovic et al. (2024)'s finding that income elasticity is economically trivial, extending their 2015-2017 result through 2022. Average marginal effects for Model 2 (reported in Table A4) confirm this: a 10% income increase reduces firewood probability by only 0.009 percentage points (AME of -0.090 for log income), nearly identical to Fercovic et al. (2024)'s finding of approximately 0.1 percentage points.

5.2 Rural Heterogeneity Analysis

Table 3 examines the differences in firewood use between rural and urban households. We continue to use model 2 as a benchmark as the rural and urban subsamples diverge from model 2 containing 2022 data. The pooled Model 2 shows a large positive Rural coefficient of 1.276 (standard error 0.012), indicating much higher firewood use probabilities in rural areas even after controlling for income, temperature, housing tenure, and forest variables. This large rural-urban gap is consistent with Fercovic et al. (2024)'s findings and persists across all three survey waves.

The linear log-income coefficient is -0.247 (standard error 0.033) in rural areas (Model 3) and -0.389 (standard error 0.015) in urban areas (Model 4), with both highly significant. This means income growth slightly reduces firewood use everywhere, but the effect is weaker in rural communes where firewood remains attractive and readily available. The rural income coefficient is approximately 37% smaller in magnitude than the urban coefficient, indicating that income-driven fuel switching is less effective in rural areas. This rural-urban heterogeneity in income elasticity extends Fercovic et al. (2024)'s main finding by showing that the economically trivial income effect is even weaker in rural areas. Forest cover and forest loss are positively associated with firewood use in both rural and urban models, with forest-loss coefficients of 0.001 (standard error 0.00003) in rural areas and 0.002 (standard error 0.00002) in urban areas. The interaction term Rural $\times$ Forest Loss in the pooled Model 2 is negative at -0.001 (standard error 0.00003), suggesting that the positive effect of forest loss on firewood use is slightly attenuated in rural areas where households may already have high baseline firewood use regardless of recent forest loss. The year dummies show that rural firewood use increased slightly more than urban use between 2015 and 2022: Year 2022 coefficient is 0.092 (standard error 0.022) in rural areas versus 0.064 (standard error 0.010) in urban areas, indicating that the rural-urban gap did not narrow over time.

Table 3: Model Progression: Rural vs Urban vs All

	Dependent variable:		
	Firewood (1 = uses firewood)		
	M2: Rural and Urban	M3: Rural	M4: Urban
	(1)	(2)	(3)
Log Total Household Income	−0.364***	−0.247***	−0.389***
	(0.013)	(0.033)	(0.015)
Log Total Household Income 2	−0.055***	−0.066***	−0.045***
	(0.010)	(0.024)	(0.012)
Log Total Household Income 3	0.067***	0.027	0.072***
	(0.015)	(0.036)	(0.017)
Minimum Temperature	−0.109***	−0.095***	−0.110***
	(0.002)	(0.007)	(0.003)
Household Size	0.103***	0.086***	0.106***
	(0.003)	(0.007)	(0.003)
Owner	0.502***	0.417***	0.513***
	(0.011)	(0.038)	(0.011)
Ceded	0.386***	0.346***	0.381***
	(0.013)	(0.040)	(0.015)
Other Ownership	0.476***	0.485***	0.463***
	(0.023)	(0.059)	(0.026)
Forest Cover	0.00000	0.00000***	0.00000***
	(0.012)	(0.00000)	(0.00000)
Forest Loss	0.002***	0.001***	0.002***
	(0.00000)	(0.00003)	(0.00002)
Log Income × Min Temperature	0.066***	0.066***	0.067***
	(0.00002)	(0.011)	(0.004)
Year = 2017	0.016*	−0.037*	0.028***
	(0.009)	(0.021)	(0.010)
Year = 2022	0.071***	0.092***	0.064***
	(0.009)	(0.022)	(0.010)
Rural	1.276***		
	(0.00003)		
Rural × Forest Loss	−0.001		
	(0.004)		
Constant	−1.369***	0.030	−1.388***
	(0.015)	(0.047)	(0.016)
Observations	155,755	34,819	120,936
Log Likelihood	−77,408.360	−13,771.850	−63,585.480
Akaike Inf. Crit.	154,848.700	27,571.700	127,199.000
Note:			*p<0.1; **p<0.05; ***p<0.01

Figure 2 visualizes predicted firewood use probabilities by income and rural status, generated from a probit model that includes Rural × Log Income interactions to allow the income-firewood relationship to differ between rural and urban areas. The model holds all other covariates at their mean values including temperature, household size, housing tenure, and forest variables. The x-axis shows log income ranging from approximately 12 to 15, corresponding to household incomes from roughly 160,000 to 3.3 million Chilean pesos. The substantial rural-urban gap persists across the entire income distribution, with rural households maintaining 75 to 85% firewood use probability even at high incomes where urban use drops from 60% to 30%. Economically, this figure demonstrates that even wealthy rural households face structural barriers to fuel switching that income growth alone cannot overcome, likely due to limited infrastructure for alternative fuels and lower opportunity costs of firewood procurement in rural areas.

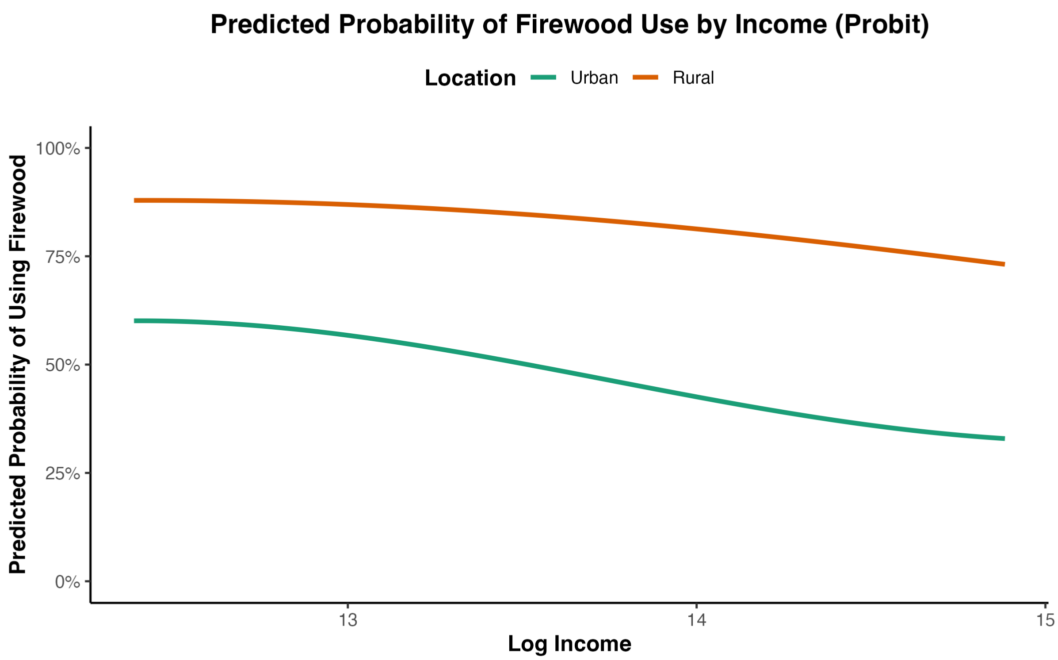

Figure 2: Predicted Probability of Firewood Use: Rural vs Urban

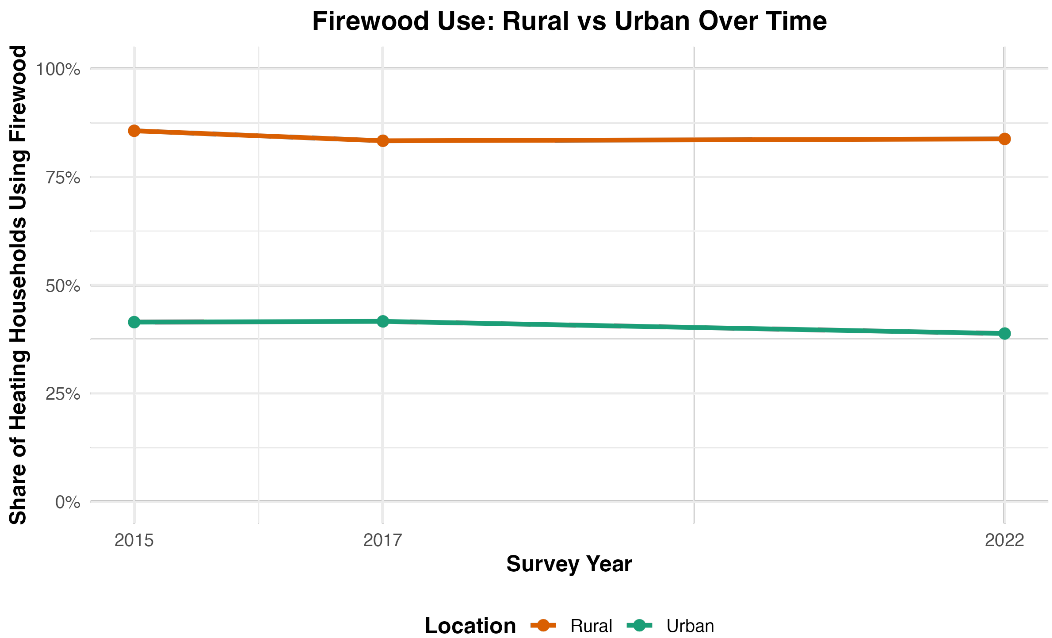

Figure 3: Firewood Use Rates: Rural vs Urban Over Time

Figure 3 plots the raw share of heating households using firewood by rural and urban location across survey years without controlling for other factors. The figure is constructed by calculating the simple mean of the firewood binary variable within each location-year combination. The persistent gap between rural and urban areas is striking. Rural firewood use remains above 70% across all years, while urban use stays below 40%. Between 2015 and 2022, rural firewood use declined only slightly from approximately 73% to 70%, while urban use remained relatively flat around 35 to 38%. Economically, this temporal stability in the rural-urban gap despite Chile's economic growth reinforces our finding that passive income-driven energy transition is insufficient, particularly in rural areas where alternative fuel infrastructure remains limited.

Figure 4 decomposes the rural-urban gap by income quintile across survey years. Households are divided into five equal-sized income groups separately for each rural-urban-year combination, and firewood use rates are calculated within each cell. This figure reveals important heterogeneity within the rural-urban divide. In rural areas, even the richest quintile maintains firewood use above 45% across all years, while in urban areas the richest quintile drops to approximately 10%. The stability of patterns across years is evident within each income group. For both rural and urban households at each income level, firewood use rates barely changed between 2015 and 2022 despite overall income growth. Economically, this demonstrates that the weak income elasticity we estimate is not driven by averaging across heterogeneous groups but rather reflects genuinely weak effects within both rural and urban subpopulations and across the income distribution.

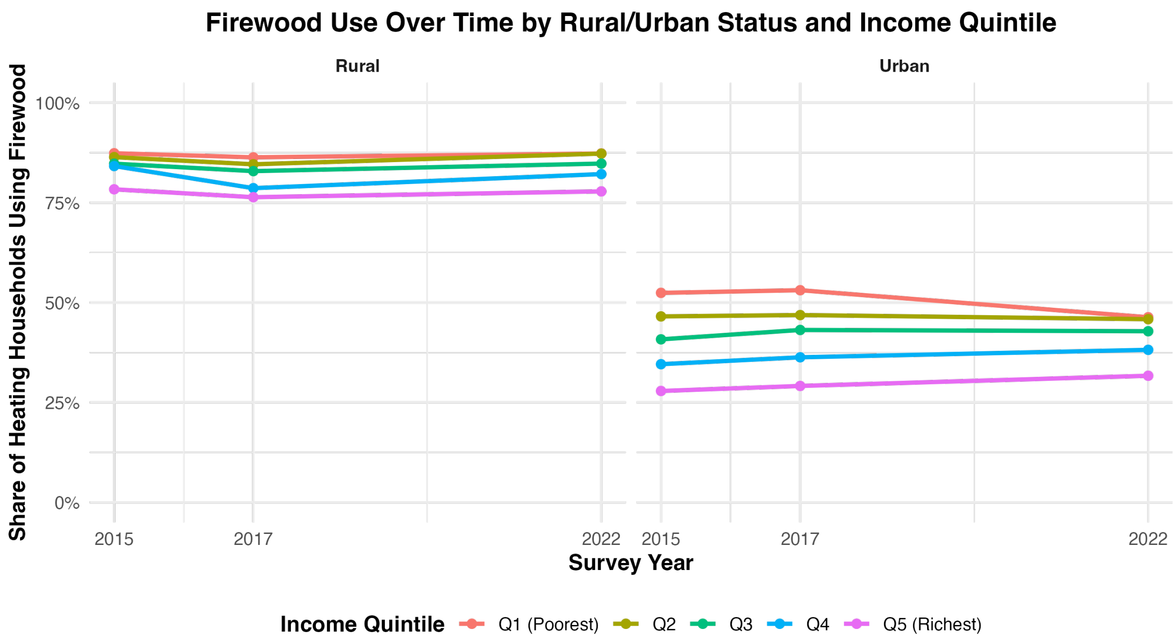

Figure 4: Firewood Use Over Time by Rural/Urban Status and Income Quintile

5.3 Heating vs Non-Heating: Selection Patterns by Rural Status

The original paper estimates a two-stage model, where households first decide whether to heat at all and then choose a fuel among heaters. Our main firewood probits condition on heating households (`Yes_Heat` = 1), but the trimmed dataset also allows us to explore the first-stage decision in a simple way while keeping the focus on rural heterogeneity.

Figure 5 plots the raw share of all households that choose to have any heating system by year and rural status, calculated as the simple mean of the heating binary variable within each location-year combination using the full sample before conditioning on heating. In all three waves, rural households are more likely to report having a heating system than urban households, with rates exceeding 96% in rural areas compared to approximately 93 to 95% in urban areas. This rural advantage in heating access remains fairly stable over time. The rural-urban gap in heating choice is smaller than the gap in firewood use among heaters, but it still matters for understanding that rural households both are more likely to heat and far more likely to choose firewood when they do heat, creating a compounding effect on firewood dependence.

Table 4 reports a probit model for heating choice in the full sample with 226,727 observations. Log income enters with a nonlinear pattern, with the linear term positive at 0.247 with a standard error of 0.005, and higher-order terms significant, indicating that higher-income households are on average more likely to heat. Minimum temperature has a small negative effect of -0.018 with a standard error of 0.002, suggesting that households in colder communes are already close to universal heating.

Other covariates behave as expected. Owners are more likely to heat than renters with an Owner coefficient of 0.231 and a standard error of 0.009. Household size reduces heating probability slightly once income and temperature are held fixed, and forest variables are positively associated with the decision to heat at all. The Rural coefficient is positive at 0.119 with a standard error of 0.010, indicating that rural households are slightly more likely to have heating systems than urban households after controlling for other factors.

Taken together, the heating-choice results show that rural households are not less likely to heat than urban households; if anything, they are slightly more likely to have a heating system. Combined with the fuel-choice results above, this implies that rural households both (i) are at least as likely to heat and (ii) are far more likely to choose firewood when they do heat. For policy, this reinforces the idea that rural areas face a double challenge: high heating prevalence and high firewood dependence.

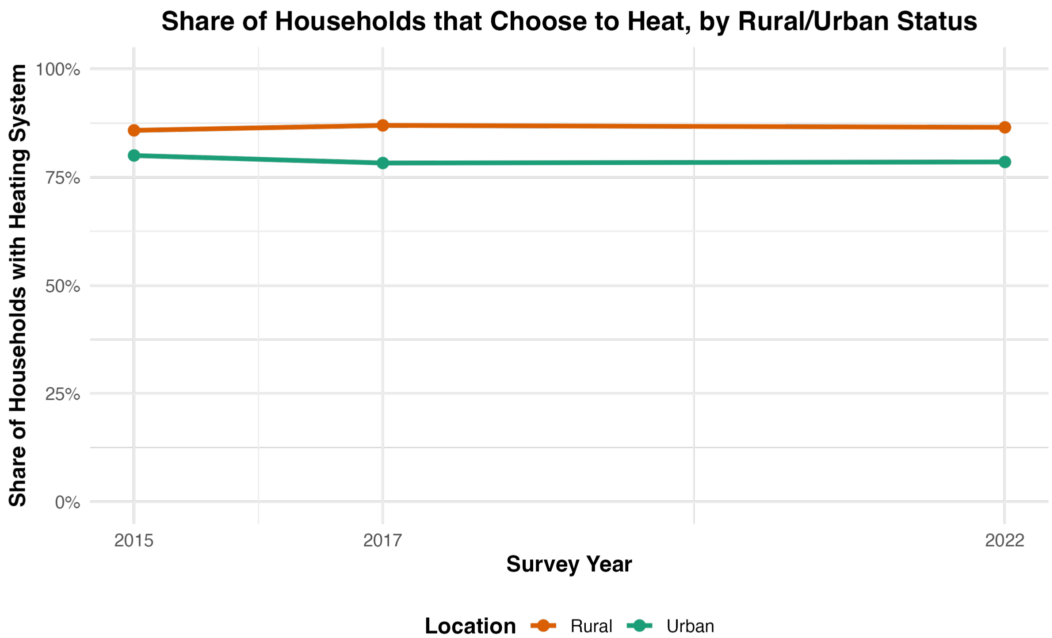

Figure 5: Share of Households that Choose to Heat, by Rural/Urban Status

Figure 6 presents predicted probabilities from the heating choice probit model, showing how the likelihood of having any heating system varies with household income for rural and urban households. The predictions are generated by varying income from the 2nd to 98th percentile while holding all other covariates including temperature, household size, housing tenure, and forest variables at their mean values. Both curves show convergence toward universal heating as income rises, with predicted probabilities approaching 99% for households earning above 2 million pesos. Economically, this figure reveals that income is indeed a strong driver of heating access itself, unlike the weak income effect on fuel choice among those who heat. The rapid rise in heating probability at low income levels suggests that budget constraints preventing any heating system installation are overcome relatively quickly with income growth. The near-convergence of rural and urban curves at higher incomes indicates that once households can afford heating, location matters less for the decision to heat but substantially more for which fuel to choose.

Table 4: Probit Model for Heating Choice (Yes_Heat = 1)

	Dependent variable:
	Has Heating System (1 = Yes)
Log_Income_c	0.247***
	(0.005)
I(Log_Income_c^2)	0.033***
	(0.003)
I(Log_Income_c^3)	0.001***
	(0.0003)
Min_Temp_c	−0.018***
	(0.002)
n_Household	−0.030***
	(0.002)
Owner	0.231***
	(0.009)
Ceded	−0.009
	(0.011)
Other_Ownership	−0.173***
	(0.018)
Rural	0.119***
	(0.010)
Forest	0.00002***
	(0.00000)
Forest_Loss	0.001***
	(0.00003)
Year	−0.014***
	(0.001)
Rural:Forest_Loss	−0.00004
	(0.00005)
Log_Income_c:Min_Temp_c	0.005***
	(0.002)
Constant	27.916***
	(2.380)
Observations	226,727
Log Likelihood	−93,865.090
Akaike Inf. Crit.	187,760.200
Note:	*p<0.1; **p<0.05; ***p<0.01

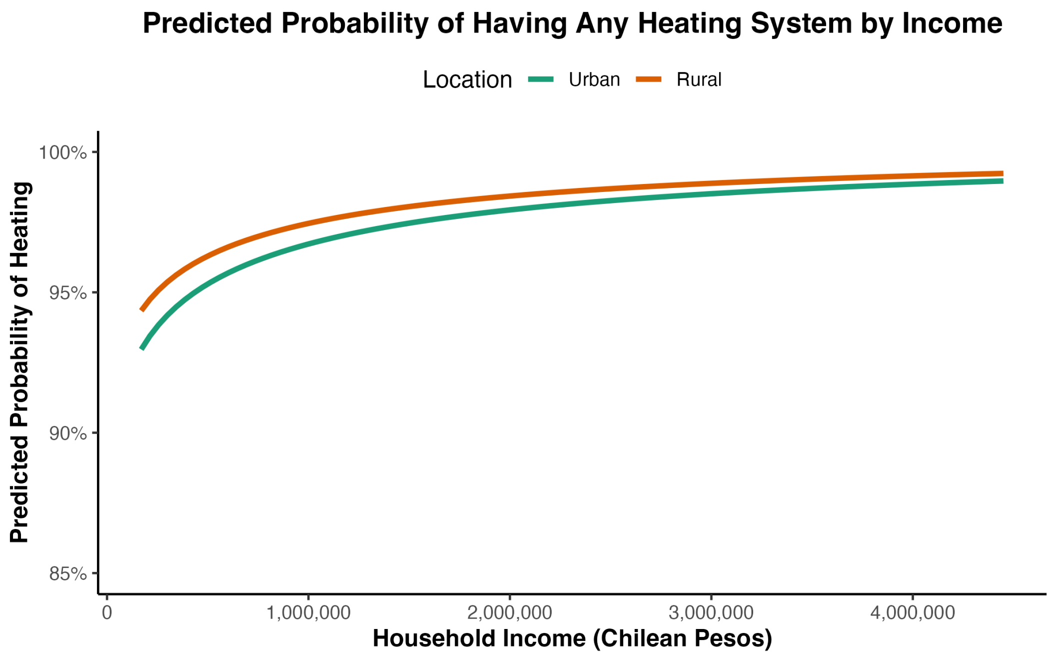

Figure 6: Predicted Probability of Having Any Heating System by Income

The stability of income elasticity across 2015-2022 provides strong evidence that Chile's high-income status creates a structural barrier to market-driven energy transition. Passive income effects remain insufficient to drive substantial fuel switching, especially in rural areas. Figure 1 shown earlier visualizes how firewood use by income quintile remained remarkably stable across 2015, 2017, and 2022. The near-parallel bars across survey years confirm that while absolute firewood use may have declined slightly, the income gradient remained essentially unchanged, consistent with our regression evidence of stable elasticity.

6 Diagnostics and Robustness

6.1 Model Diagnostics

We conduct comprehensive diagnostics to assess model reliability and validate our main findings. Our diagnostic strategy follows Assignment 3 methods but extends them to account for the temporal dimension and multiple model specifications in our final project.

First, we calculate Variance Inflation Factors (VIF) for all regressors to check for multicollinearity using a linear probability model equivalent, as probit models do not directly support VIF computation. Polynomial income terms are expected to show high VIF mechanically, but mean-centering substantially mitigates this issue. We use a threshold where VIF less than 10 is acceptable, with values below 5 considered excellent. Our VIF analysis confirms that multicollinearity is not a concern. The linear income term shows VIF of 4.58 and the cubic term 4.52, both acceptable. Non-polynomial main effects show even better values: Rural (1.68), Min_Temp_c (1.37), n_Household (1.15), Forest (1.34), and Forest_Loss (1.50). The Rural × Forest_Loss interaction shows VIF of 2.15, and the Log_Income × Min_Temp interaction shows VIF of only 1.06. All values are well within acceptable bounds, indicating that our polynomial specification with centering successfully avoids multicollinearity issues that could inflate standard errors or destabilize coefficient estimates.

Second, we test for heteroskedasticity using Breusch-Pagan (BP) tests. Since BP tests are designed for linear models, we estimate Linear Probability Models (LPM) as approximations of our probit specifications for models 1 through 4. Table A7 reports BP test statistics, p-values, and degrees of freedom for each model. The BP test results indicate that heteroskedasticity is detected in all four models (p-values less than 0.05), which is expected in cross-sectional household data.

For model fit, we report McFadden pseudo-R^2, log likelihood, and AIC for all specifications. The McFadden pseudo-R^2 values range from 0.087 for the rural-only model (M3) to 0.283 for the full pooled model (M2), indicating reasonable fit given the binary outcome and cross-sectional nature of the data. The rural-only model's lower pseudo-R^2 of 0.087 reflects less variation to explain once we condition on rural location, as rural households show more homogeneous firewood dependence regardless of income or other covariates. Model 1 achieves a pseudo-R^2 of 0.279 with 105,874 observations from 2015 and 2017. Model 2 with all three waves achieves pseudo-R^2 of 0.283 with 155,755 observations and log likelihood of $-77,408.5$.

The AIC for Model 2 is 154,847 compared to Model 1's 105,755. The supplementaryheating choice model achieves pseudo-R^2 of 0.161 with 226,727 observations.

We assess classification accuracy at the 0.5 probability threshold for all models using confusion matrices. The results show strong predictive performance. The baseline Model 1 correctly classifies 77.0% of observations. The pooled Model 2 achieves similar accuracy at 77.3%. The rural-only model (M3) classifies 84.4% correctly, while the urban-only model (M4) achieves 75.3% accuracy. The notably higher accuracy in the rural subsample reflects the fact that rural households have more homogeneous and predictable firewood use patterns, with the vast majority using firewood regardless of income. The high accuracy rates confirm that our models capture the main determinants of firewood use effectively.

We examine fitted versus actual probability plots for the main pooled Model 2 by plotting predicted probabilities against observed binary outcomes. Figure 7 shows the relationship between predicted and actual firewood use, with the points showing the binary nature of the outcome variable. This pattern is consistent with proper probit specification and indicates no systematic prediction failure or functional form misspecification.

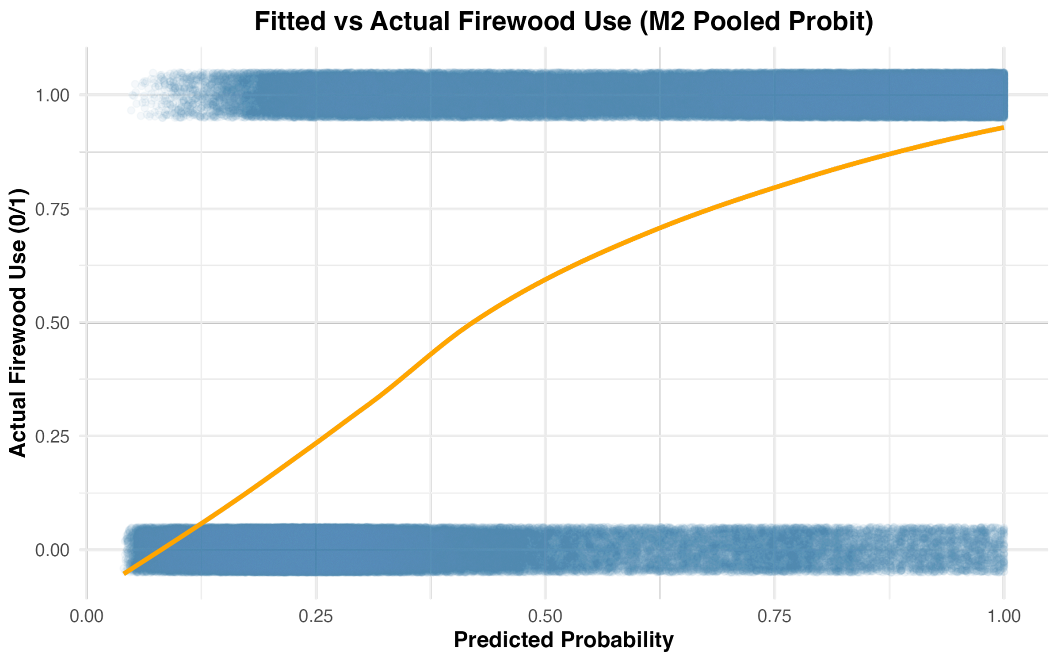

Figure 7: Fitted vs Actual Firewood Use: Pooled Model M2 (2015-2022)

To check for influential observations that survived outlier trimming, we compute Cook's distance for the pooled Model 2 and identify observations exceeding the threshold of 4 divided by sample size, which equals approximately 0.000026. Figure 8 plots Cook's distance for all 155,755 observations, with the red dashed line indicating the threshold and red points highlighting potentially influential observations. Using this criterion, approximately 2.3% of observations are flagged as potentially influential. The vast majority of observations show Cook's distance well below the threshold, with only a small fraction of points appearing above the red line. These influential observations are distributed across communes and years without evidence of data errors or geographic clustering. The presence of some high Cook's distance values is expected in large samples and does not indicate model misspecification.

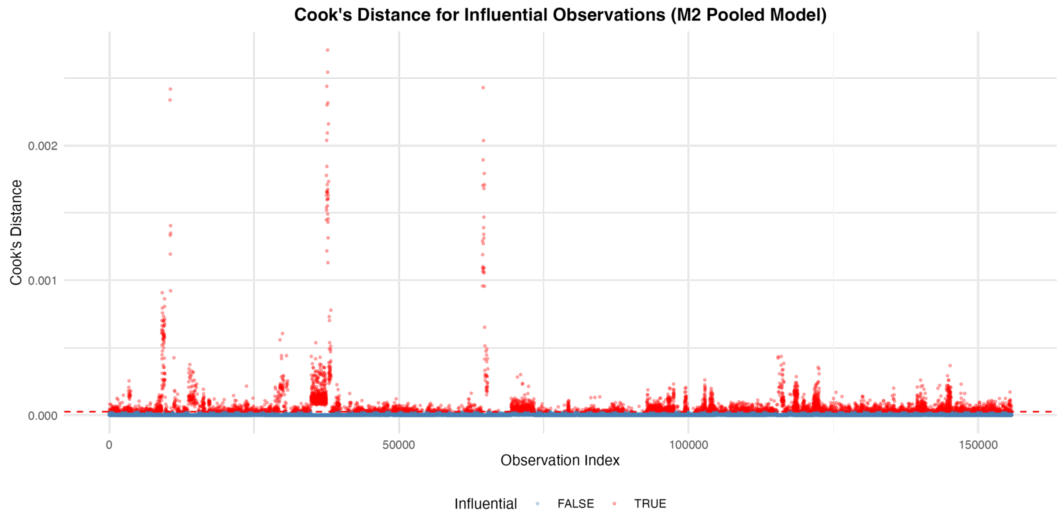

Figure 8: Cook's Distance for Influential Observations: Pooled Model M2

6.2 Robustness

We conduct several robustness checks to validate our main findings and demonstrate that our conclusions are not sensitive to modeling choices.

Based on provided feedback, we tested alternative functional forms to assess whether our cubic specification is justified. We estimate a model without the cubic income term, retaining only linear and quadratic terms. This restricted specification shows a modestly worse AIC of 154,864 compared to the full model's 154,847, an increase of 17 points. The pseudo-R^2 remains 0.283, identical to the full model. The income effect becomes more linear but remains economically small, confirming that our finding of trivial income elasticity does not depend critically on the cubic specification. We also estimate a model without interaction terms, dropping both Rural × Forest_Loss and Log_Income × Min_Temp. This simplified model shows substantially worse fit with AIC of 156,361 compared to 154,847 for the full model, an increase of 1,514 points. The pseudo-R^2 declines from 0.283 to 0.276, indicating that interactions capture meaningful heterogeneity in how forest loss and temperature affect different groups.

Although we primarily report probit results for consistency with Fercovic et al. (2024), we verify that our conclusions are not sensitive to the choice between probit and logit link functions. We re-estimate all specifications using logit and compare marginal effects. The average marginal effect of log income differs by less than 0.003 percentage points between probit and logit across all specifications. Predicted probabilities from probit and logit models are nearly identical across the income distribution, with correlation exceeding 0.998. This confirms that functional form choice between these two standard binary models does not materially affect our conclusions.

7 Policy and Industry Implications

7.1 Implications for Chilean Energy Policy

Our finding of stable and trivial income elasticity across 2015-2022 has clear policy implications. Passive policies relying on economic growth to solve the firewood dependence problem are empirically rejected by our temporal validation.

Several interventions are necessary to achieve meaningful energy transition. Direct financial support for infrastructure should specifically target rural households. The persistent rural-urban gap with 40 to 60% firewood use in rural areas versus less than 20% in urban areas indicates the need for gas network expansion and clean energy access in rural areas.

Energy transition requires fiscal commitment. Our evidence that income elasticity did not strengthen by 2022 suggests threshold effects where higher income suddenly triggers switching do not apply in Chile's high-income context. Policymakers cannot wait for households to naturally transition as incomes rise.

8 Conclusion

We replicated Fercovic et al. (2024)'s finding of economically trivial income elasticity. Our baseline probit estimates show that a 10% income increase reduces the probability of using firewood by only a few hundredths of a percentage point, consistent with the original paper's finding of about 0.1 percentage points, both statistically significant but economically trivial. We extended the analysis to include the 2022 CASEN wave, creating a three-wave pooled cross-section with a total of 199,955 observations across the three years. The income elasticity remained stable across 2015-2022. Rural heterogeneity persisted as rural households maintain 40 to 60% firewood use probability even at high incomes compared to urban households with less than 20%. Forest availability effects remained positive and significant, and we identified CASEN as a repeated cross-section rather than a true panel, necessitating pooled cross-sectional methods instead of panel estimators.

Our temporal extension provides critical validation evidence. Fercovic et al. (2024)'s pessimistic conclusion about income-driven energy transition is not period-specific but rather a stable feature of Chile's context. Five years of additional economic development from 2017 to 2022 did not materially alter the trivial income elasticity, strengthening the case for direct policy intervention rather than passive reliance on economic growth. The stability of the findings across 2015-2022 increases our confidence that observed patterns reflect structural relationships rather than temporary phenomena.

Several limitations merit discussion. CASEN uses rotating samples, preventing household-level fixed effects estimation, so we cannot control for unobserved household heterogeneity. However, stability across three independent cross-sections provides robustness against period-specific shocks. We use 2019 Hansen forest data merged to 2022, treating forest cover as time-invariant proxy. This introduces measurement error for 2022 observations. Minimum temperature is treated as time-invariant based on high 2015-2017 correlation exceeding 0.99, though long-term climate change effects are not captured. Unlike Fercovic et al. (2024)'s two-stage approach with heating choice followed by fuel choice, we condition directly on heating households for computational simplicity, which may introduce selection bias if confounding factors affecting heating choice also influences fuel choice. Income is potentially endogenous due to omitted variable bias, so without instrumental variables, our elasticity estimates are less credible. The 2022 data captures post-pandemic recovery, potentially biasing estimates with crisis-induced behavioral changes. Forest availability proxies firewood supply but not prices, so regional firewood price variation could confound forest effects.

Future research should extend the analysis with additional CASEN waves to test whether elasticity evolves non-linearly and conduct experimental evaluation of subsidy programs to estimate causal price elasticities. It's also worth considering the possibility of non-linear income effects and the role of other factors such as social norms and cultural preferences.

References

Appendix: Diagnostic Tables

This appendix presents detailed diagnostic statistics referenced in Section 6.

A1. Breusch-Pagan Tests for Heteroskedasticity

Table A1 reports Breusch-Pagan test results for heteroskedasticity using Linear Probability Model approximations of probit models 1 through 4. All four models show statistically significant evidence of heteroskedasticity (p-values less than 0.05).

Table 5: Breusch-Pagan Tests for Heteroskedasticity (Linear Probability Models)

Model	BP.Statistic	p.value	df	Significant
M1: Baseline (2015+2017)	4,298.090	0	14	Yes
M2: Pooled (2015-2022)	6,534.534	0	15	Yes
M3: Rural Only	2,186.676	0	13	Yes
M4: Urban Only	4,402.562	0	13	Yes

A2. Variance Inflation Factors

Table A2 reports VIF values for all variables in the pooled Model 2. All values are well below the threshold of 10, with main effects showing VIF below 5, confirming that multicollinearity is not a concern.

Table 6: Variance Inflation Factors - Pooled Model

Variable	VIF
Log_Income_c	4.580
I(Log_Income_c^2)	1.170
I(Log_Income_c^3)	4.520
Min_Temp_c	1.370
n_Household	1.150
Owner	1.800
Ceded	1.720
Other_Ownership	1.160
Rural	1.680
Forest	1.340
Forest_Loss	1.500
Year	1.130
Rural:Forest_Loss	2.150
Log_Income_c:Min_Temp_c	1.060

A3. Model Fit Comparison

Table A3 compares fit statistics across all six model specifications. The pooled model (M2) achieves the highest pseudo-R^2 values around 0.28, while the rural-only model (M3) shows lower pseudo-R^2 but higher classification accuracy due to homogeneous firewood use in rural areas.

Table 7: Model Fit Statistics Comparison

Model	Observations	McFadden_R2	Log_Likelihood	AIC	Classification_Accuracy
M1: Baseline (2015+2017)	105,874	0.279	-52,862.700	105,755.400	0.770
M2: Pooled (2015-2022)	155,755	0.283	-77,408.400	154,848.700	0.773
M3: Rural Only	34,819	0.088	-13,771.800	27,571.700	0.844
M4: Urban Only	120,936	0.222	-63,585.500	127,199	0.753
M_Heat: Heating Choice	226,727	0.161	-93,865.100	187,760.200	

A4. Average Marginal Effects

Table A4 reports average marginal effects for the pooled Model 2. The income effect of -0.090 (standard error 0.002) implies a 10% income increase reduces firewood probability by 0.009 percentage points, confirming economic triviality despite statistical significance. This matches Fercovic et al. (2024)'s finding of approximately 0.1 percentage points for a 10% income increase.

Table 8: Average Marginal Effects - Pooled Model M2 (2015-2022)

Variable	AME	SE	z value	p value
Forest Loss (km²)	0.0005	0	159.876	0
Log Income (centered)	-0.090	0.002	-40.543	0
Min Temperature (centered)	-0.030	0.001	-43.784	0
Household Size	0.029	0.001	38.379	0
Rural (1=Yes)	0.309	0.002	130.982	0

A5. Rural vs Urban Marginal Effects

Table A5 compares marginal effects between rural and urban subsamples. Income effects are stronger in urban areas (-0.101) than rural areas (-0.049), but both remain economically trivial. A 10% income increase reduces firewood probability by 0.0101 pp in urban areas and 0.0049 pp in rural areas.

Table 9: Marginal Effects Comparison: Rural vs Urban Subsamples

Subsample	Variable	AME	SE	p value
Rural	Forest Loss (km²)	0.0002	0.00001	0
Rural	Log Income (centered)	-0.050	0.004	0
Urban	Forest Loss (km²)	0.001	0	0
Urban	Log Income (centered)	-0.101	0.003	0

A6. Alternative Specification Comparison

Table A6 compares the full model against specifications without the cubic term and without interactions. Dropping interactions increases AIC by 1,514 points, indicating interactions capture meaningful heterogeneity. Dropping the cubic term increases AIC by only 17 points, but we retain it for theoretical flexibility.

Table 10: Alternative Specification Comparison

Specification	Observations	McFadden_R2	AIC	BIC
Full Model (M2)	155, 755	0.283	154, 848.700	155, 008
Without Cubic	155, 755	0.283	154, 864.400	155, 003.800
Without Interactions	155, 755	0.276	156, 360.900	156, 490.300

AI Use Appendix