Logit Choice Model with Applications

Meredith Fowlie

April 2021

Outline for today

- Welfare measures and the conditional logit model.
- Application 1: Davis (2021)
- Application 2: Burgess et al. (2020)
- Introducing .... Mixed Logit!

Recall our choice context (in general terms)

Remember the kinds of choice situations are we thinking about-

- The dependent variable, Y, takes on non-negative, un-ordered interger values between zero and J.
- Typically motivated as an economic choice (i.e. some kind of constrained optimization) from a set of discrete options or alternatives.
- Today we'll focus exclusively on random utility maximization (RUM).

Conditional logit model

Basic set-up:

$$\mathrm{U}^*_{nj} = U(X_{nj}; \beta) + \varepsilon_{nj}$$
$$\varepsilon_{nj} \sim iidEV1$$

- Covariates X_{nj} vary across choices. Choices can also vary by indivudual.
- Unobserved component assumed to be $\sim$ iid distributed according to the extreme value distribution (similar to normal.. but fatter tails and asymmetric).
- Why the extreme value distribution? It yields analytically integrable conditional choice probabilities!

From choice modeling to welfare analysis

- We are often interested in estimating the welfare impacts of a change in choice sets, changes in attributes, price changes.
- A primary reason to estimate discrete choice models: If you can estimate the parameters of RUM, you can derive consumer surplus changes associated with changes to the choice set/context.
- The use of logit models derived from RUM supports/expands possibilities for substantive welfare analysis.

Discrete choice meets welfare analysis

Conditional logit model is derived from a theoretical model of latent utility maximization that can be used to quantify welfare impacts:

1. Partial/marginal welfare analysis can also be supported under weaker assumptions.
2. Leaning on structural assumptions, we can in principle estimate non-marginal changes in consumer surplus associated with changes to the choice set.

Partial welfare analysis

Start with a quasilinear utility function (z is consumption of numeraire with price normalized to 1):

$$max_{j,z}\ u_j + z\ s.t.\ p_j + z = y$$

Making the substitution

$$max_{j,z}\ u_j + y - p_j$$

Suppose latent utility takes the form:

$$u_j = \beta' X_j - \alpha p_j$$

This assumes away income effects.

Partial welfare analysis using WTP

$$u_j = \beta' X_j - \alpha p_j$$

Taking the total derivative wrt changes in choice attribute X and price P:

$$du_j = \beta dX_j - \alpha dp_j$$

Set this to zero to find the change in the price that keeps the individual just indifferent to a change in the attribute:

$$WTP = \frac{dp_j}{dX_j} = \frac{\beta}{\alpha}$$

Discrete changes in consumer surplus and discrete choice

- Random utility maximization (RUM): Consumers are assumed to have consistent and transitive preferences over discrete alternatives.
- In a discrete choice context, direct integration under a continuous demand curve is not possible.
- Consumer surplus can be defined as the utility associated with the utility maximizing choice.

Consumer surplus measures in a discrete choice framework

- Consumer surplus is the utility (monetized) that a person receives in a choice situation.
- If a consumer's marginal utility of income α_n is constant over the price change, we can translate utility into \$:

$$CS_n = \frac{1}{\alpha_n} Max_j(U^*_{nj} \quad \forall \quad j)$$

Division by α translates utility into dollars.

Random Utility Maximization

Recall our decomposition of the latent utility:

$$U^*_{ni} = \underbrace{U(X_{ni};\beta)}_{\text{deterministic component}} + \underbrace{\epsilon_{ni}}_{\text{unobserved component}}$$

We do not observe the stochastic component ϵ.

With this random component, RUM takes on a probabilistic form.

Consumer surplus and the logit model

If we can assume the following:

1. The unobserved component of utility is independently and identically distributed EV1 (logit)
2. Utility is linear in income

Then the expected consumer surplus has a closed form.

Small and Rosen(1981) show that the expected utility derived from a choice set characterized by X_n is a function of the log of the denominator of a logit choice probability (a.k.a. the log sum):

$$CS_n = \frac{1}{\alpha_n} ln(\sum_j e^{\beta_n' X_{nj}}) + C \tag{1}$$

The Logsum

- Expected consumer surplus in the logit model is the log of the denominator of the logit choice probability (divided by α and adding).
- $ln(\sum_j e^{\beta_n' X_{nj}})$ is often called the *log sum* term.
- Because the level of utility is not observed, the constant C is not identified.

Welfare impacts of a price/policy change

Good news.. we are typically interested in evaluating the change in consumer surplus that results from a change in choice attributes/choice: *changes*:

$$\Delta E[CS_n] = \frac{1}{\alpha_i}[ln(\sum_j \exp(U^0_{nj})) - ln(\sum_j \exp(U^1_{nj}))]$$

- Note that the unidentified C parameter drops out of this difference.
- Assumption of constant marginal utility of income might be more tenable over the range of potential income changes implicated by the policy or market change of interest.

Income effects??

- This standard logsum approach is based on quasi-linear preferences (no income effects - marginal utility of income is assumed constant)
- A price or cost variable enters the representative (indirect) utility in a consistent linear additive fashion. The negative of its coefficient is the marginal utility of income by definition.
- In cases where income effects are important OR in cases where income enters in a non-linear way, the derivation of consumer surplus measures are more complicated.

Consumer surplus and the logit model

- $E[CS_n]$ is the average consumer surplus for the subpopulation of agents with the same utility function as person n.
- One approach to calculating total consumer surplus is to take a population weighted sum of representative average CS estimates.
- In the logit case, total consumer surplus in the population can be calculated as a weighted sum of logsums over a sample of decision-makers. with the weights reflecting the number of people in the population who face the same representative utilities as the sampled person.

Outline for today

- Welfare measures and the conditional logit model.
- Application: Davis (2021)
- Application 2: Burgess et al. (2020)
- Introducing Mixed Logit

What Matters for Electrification?

- The share of households heated with electricity has risen from 8 percent in 1970 to 39 percent in 2018.
- Lucas wants to analyze this discrete choice of heating fuel using household-level data.
- He first estimates a linear probability model which helps him identify key determinants of this choice.
- He then uses a conditional logit RUM framework to assess the welfare implications of an electrification mandate for new homes.

Davis(2021)

- The most promising path to decarbonization is electrification! But how do we get there?
- Critics argue that electrification mandates can be expensive and regressive.
- Proponents argue it's the best way to make a needed transition.

Empirical context

- Lucas uses data on heating choices from millions of households over a 70 year period to investigate key determinants of this discrete heating choice.
- US Census and the Americal Community Survey ask respondents about primary form of home heating.
- Residential fuel prices constructed as annual revenues/sales.
- Other covariates include heating degree days (HDD) and household/house characteristics (e.g. income, number of bedrooms).

Linear Probability Model

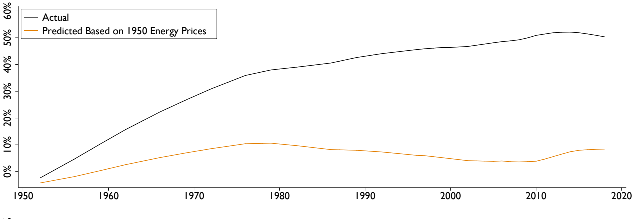
- Dependent variable is an indicator for electric heating.
- Predictors include: electricity price, gas price, HDD, income, house size, etc.
- Focus exclusively on new builds (that's when the heating fuel choice is primarily made).
- Coefficients estimate how an increase in X increases the probability of electric heating.
- These linear regressions used to decompose the total increase in electric heat adoption.

Table 2: Linear Probability Model, Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
Electricity Price, in logs	-0.40**	-0.43**	-0.38**	-0.40**	-0.40**	-0.42**
	(0.03)	(0.04)	(0.03)	(0.05)	(0.04)	(0.06)
Natural Gas Price, in logs	0.21**	0.29**	0.18**	0.24**	0.15**	0.21**
	(0.06)	(0.08)	(0.05)	(0.07)	(0.05)	(0.07)
Heating Oil Price, in logs	0.04	-0.08	0.08*	0.08	0.09**	0.06
	(0.04)	(0.15)	(0.03)	(0.10)	(0.03)	(0.10)
Household Income, 1000s	-0.00**	-0.00**	-0.00**	-0.00**	-0.00**	-0.00**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Heating Degree Days, 1000s	-0.06**	-0.06**	-0.09**	-0.04*	-0.05**	-0.06**
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
Four Bedroom Home	-0.05**	-0.05**	-0.05**	-0.04**	-0.05**	-0.05**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Five+ Bedroom Home	-0.10**	-0.08**	-0.10**	-0.08**	-0.10**	-0.08**
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)
Rented, i.e. not owned	0.01	0.01	0.01	0.02*	0.02*	0.02**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Mobile Home	0.04	0.02	0.03	0.02	0.03	0.02
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Single Family, Attached	0.04*	0.04**	0.04**	0.04**	0.04**	0.03**
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Multi-Unit Home, 2-4 Units	0.12**	0.12**	0.13**	0.12**	0.12**	0.12**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Multi-Unit Home, 5+ Units	0.25**	0.24**	0.25**	0.24**	0.25**	0.24**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Year Fixed Effects	No	Yes	No	Yes	No	Yes
Geographic Fixed Effects	No	No	Regions	Regions	Divisions	Divisions

Key driver of electric heat adoption is fuel prices

Figure 5: Percentage of New Homes Heated with Electricity, Decomposition



Limits of linear probability models

- Linear probability models provide a relatively simple way to predict choices conditional on choice attributes.
- But very limited in what you can say about how choices or welfare might be impacted under counterfactual choice contexts.
- Lucas wants to know: How much income would households need to receive to make them indifferent between status quo and an electrification mandate?
- He uses the simplest possible CL model to answer this question.

Random utility model

Households choose heating to maximize:

$$u_{ij} = \overset{\text{FE}}{\alpha_{0j}} + \alpha_1 \overset{\text{p based on location/climate}}{p_{ij}} + \alpha_2 x_i + \epsilon_{ij}$$

- Here j denotes one of two heating choices: electric (e) or gas (g).
- He drops from the sample less common fuels (e.g. propane).
- The deterministic component of latent utility depends on annual heating costs p and housing/climate characteristics x.
- How do we identify α_2?

We can't! Only relative differences matter!

We have to normalize the model in more ways than one!

- Normalize by unidentified scale parameter to pin down the scale of the model.
- Define gas as the baseline choice. The α_{0e} and α_{2e} coefficients capture differences relative to gas.

$$u_{ie} = \alpha_{0e} + \alpha_1 p_{ie} + \alpha_{2e} x_i + \epsilon_{ie}$$
$$u_{ig} = \alpha_1 p_{ig} + \epsilon_{ig}$$

change in heating price
regardless of source

How to interpret α_{0e}?

We can't! Only relative differences matter!

We have to normalize the model in more ways than one!

- Normalize by unidentified scale parameter to pin down the scale of the model.
- Define gas as the baseline choice. The α_{0e} and α_{2e} coefficients capture differences relative to gas.

$$
\begin{aligned}
u_{ie} &= \alpha_{0e} + \alpha_1 p_{ie} + \alpha_{2e} x_i + \epsilon_{ie} \\
u_{ig} &= \alpha_1 p_{ig} + \epsilon_{ig}
\end{aligned}
$$

How to interpret α_{0e}?
It captures the average difference in latent utility relative to gas, conditional on other variables in the model, scaled by the unobserved σ.

Table 5: Heating System Choice Model

	Estimated Coefficients	Implied Marginal Effects
Annual Energy Expenditures, in 1000s	-1.40**	-0.35**
	(0.31)	(0.08)
Electric Heating System x		
Household Income, 100,000s	-0.18**	-0.04**
	(0.03)	(0.01)
Heating Degree Days, 1000s	-0.21*	-0.05*
	(0.09)	(0.02)
Four Bedroom Home	-0.43**	-0.11**
	(0.04)	(0.01)
Five+ Bedroom Home	-0.64**	-0.16**
	(0.11)	(0.03)
Rented, i.e. not owner-occupied	0.45**	0.11**
	(0.05)	(0.01)
Mobile Home	1.42**	0.35**
	(0.17)	(0.04)
Single Family Home, Attached	-0.28*	0.07*
	(0.12)	(0.03)
Multi-Unit Home, 2-4 Units	0.50**	0.12**
	(0.09)	(0.02)
Multi-Unit Home, 5+ Units	1.09**	0.27**
	(0.12)	(0.03)
Constant	1.84	–
	(0.53)	

Note: This table reports coefficient estimates and standard errors as well as marginal effects and standard errors from a conditional logit model estimated using maximum likelihood with data on heating system choices from 950,469 households. The estimation sample includes all homes that are heated with electricity or natural gas and under ten years old in the American Community Survey samples 2000,

Interpretation?

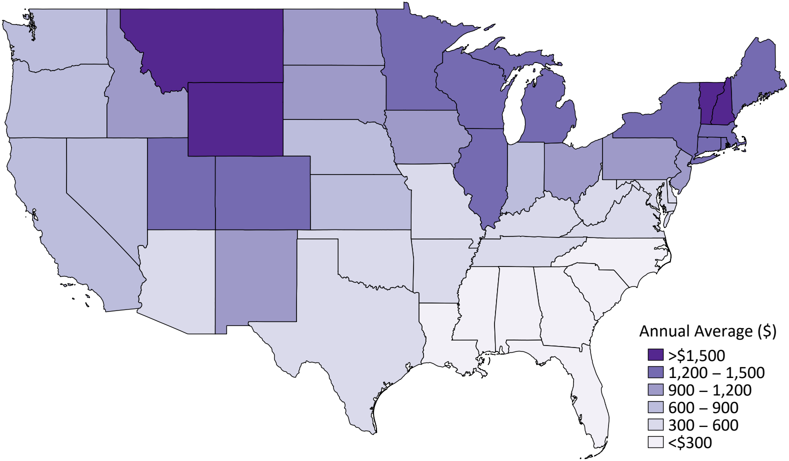
- Estimated parameters difficult to interpret.
- Marginal effects more intuitive: A $1000 increase in annual expenditures decreases the probability that a household selects that alternative by 35 percentage points.
- Interaction terms very difficult to interpret as we discussed. But can improve the choice probabilities by allowing the preference for electric versus gas heat to vary with house/climate attributes.

Welfare impact of an electrification mandate?

$$WTP_i = \frac{1}{|\alpha_1|}[\ln(e^{\alpha_{0e}+\alpha_1 x_{ie}+\alpha_{2e}z_i} + e^{\alpha_1 x_{ig}}) - \ln(e^{\alpha_{0e}+\alpha_1 x_{ie}+\alpha_{2e}z_i})]. \tag{5}$$

Dividing by the marginal utility of income α_1 translates utility into dollars. In addition, willingness-to-pay depends on energy expenditures (x_{ij}), household characteristics (z_i), and the other model parameters α_{0e} and α_{2e}. Households who strongly prefer natural gas have a high willingness-to-pay while households who strongly prefer electricity have willingness-to-pay near zero.

Figure 7: Willingness-to-Pay to Avoid an Electrification Mandate



Take-aways?

- Paper documents a dramatic increase in electric heating.
- Energy prices explain over 70 percent of this increase.
- Households in warm climates are close to indifferent between electric and gas.
- Households in cold states prefer natural gas. Would be made worse off by $2000 (or more) annually.
- Comments?

Outline for today

- Welfare measures and the conditional logit model.
- Application: Davis (2021)
- Application 2: Burgess et al. (2020)
- Introducing Mixed Logit

Energy Demand on the Electrification Frontier

- In pursuit of economic growth, many countries make major capital investments in expanding access to electricity.
- Rapid decline in solar PV costs has changed the shape of this frontier.
- Households now have a choice between energy access alternatives: grid electricity, micro-grid, off-grid solar, diesel.

Demand for Electricity on the Global Electrification Frontier

Research objectives?

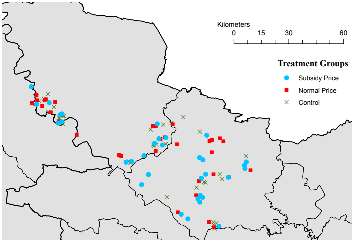
- Study household demand for electrification using a revealed-preference measure of the value of electricity.
- Estimate how households value both grid and off-grid alternatives within a single demand system (so as to study substitution between sources).
- Use experimental variation in electricity price to address endogeneity concerns.

  "This study joins a methodological movement in the development literature that combines structural models with experimental variation to aid in the interpretation and increase the validity of experimental results."

Field experimentation meets (nested) logit

- Authors implement a large-scale field experiment which randomly varied the offered tariff for solar microgrids across villages in Bihar, India.
- They track households' choice of electricity supply sources before/during/after the experiment over the period 2013-2017.

A Study districts within the state of Bihar, India

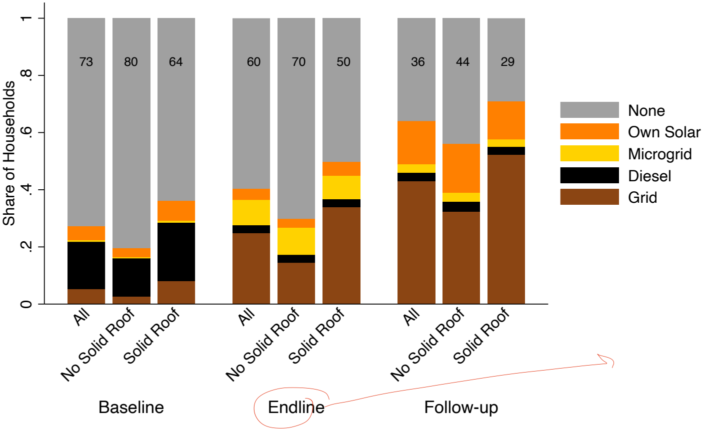


B Sample villages within study districts

The figure shows the study area. Panel A highlights the two districts of West Champaran and East Champaran, in the northwest corner of Bihar, which contain the study villages. Panel B shows, within the two study districts, the locations of sample villages and their treatment assignments. The nearest large towns are Bettiah and Motihari. The river Gandak, in the northwest, forms the state border with Uttar Pradesh.

Electricity access is a discrete choice in Bihar

- Experimental estimates of microgrid demand provide an incomplete measure of the demand for electricity access because there are several alternatives.
- Plummeting solar prices, accelerated electrification efforts, and reliability improvements over this period also mean that households' choice set changed significantly between baseline and endline.
- These authors nest the microgrid experiment within a larger demand analysis to estimate how households value both grid and off-grid alternatives within a single demand system.



The figure shows the market shares of different sources of electricity over time. Each stacked bar gives the share of households, from bottom to top, that use grid electricity, diesel generators, solar microgrids, own solar systems or no electricity. These market shares are calculated with respect to the total sample of households, without regard for whether a source is available in a village or not; in a village where the grid is not present, for example, the grid necessarily has a zero share. There are three clusters of bars, for shares in the baseline (starting November 2013), endline (starting May 2016) and follow-up (starting May 2017) survey waves. We use a dummy variable for whether a household has a solid roof as a proxy for household assets. Within each cluster of bars, the three bars from left to right give the market shares amongst all households, households that do not have a solid roof, and households that do have a solid roof, respectively.

Table 2: Description of Electricity Sources in Bihar

	Grid electricity (1)	Microgrid solar (2)	Own solar (3)	Diesel generator (4)	No electricity (use kerosene) (5)
Availability	Grid must reach village. Then household can apply for a connection.	Offered in treatment villages by Husk Power Systems.	Available in market towns. Households travel to buy on their own.	Private operators offer in villages with high enough demand.	Sold through Public Distribution System.
Energy services typically supported	Light and phone charging, fans, televisions. No load limit.	Light and phone charging.	Light and phone charging. Fans for larger systems.	Light and phone charging.	Light, of lower quality.
Reliability	Poor. Frequent power cuts at peak times.	Good, but limited by battery capacity.	Good, but limited by battery capacity.	Good, but operates during peak evening hours only.	Good.
Contract	Pay monthly bill, either flat or per unit.	Pay monthly flat bill.	Buy system up-front. Low marginal cost but household liable for maintenance.	Pay monthly flat bill.	Pay by quantity at a subsidized rate.
Risk of disconnection	Disconnection possible, though unlikely, if no payment.	Disconnection possible if no payment.	Not applicable.	Disconnection likely if no payment.	Not applicable.

The table describes the electricity sources that are used by households in our sample in Bihar.

Energy transformation take-aways

1. Electrification rate is surging over this time period in Bihar.
2. Diesel is in decline.
3. Richer households more likely to have grid connections.
4. Off-grid solar more common among poorer households.

Experimental price variation

The authors generate experimental variation across villages in the microgrid price with a cluster-randomized control trial.

Table 4: Solar Microgrid Demand by Village Treatment Arm

	Survey		
	Baseline	Endline	Follow-up
	(1)	(2)	(3)
Treatment: Subsidized price	-0.001	0.193***	0.081***
	(0.005)	(0.049)	(0.027)
Treatment: Normal price	0.009	0.060**	0.020*
	(0.010)	(0.028)	(0.012)
Constant	0.006	0.023***	0.002
	(0.004)	(0.005)	(0.002)
Observations	100	100	100

The table shows estimates of microgrid demand by treatment status. The dependent variable is the village-level market share of microgrid solar from survey data, which measures whether households report having the source. There are three treatment arms: a subsidized price arm (microgrids offered at INR 100), a normal price arm (microgrids offered at the prevailing price of INR 200, later cut to INR 160 in some villages), and a control arm (microgrids not offered). Each

- Subsidies increase take up of micro-grid.
- Microgrid shares fell sharply by follow-up survey.

Nice field experiment. Why not end there?

- Estimated demand for this one energy source provides an incomplete welfare picture because there are other close substitutes available.
- The elasticity of demand for microgrids cannot be used to measure the willingness to pay for electricity access. When microgrid price rises, households may substitute to other options.
- To evaluate the welfare gains from electricity access, the authors specify and estimate a demand model that incorporates all access options.

Discrete choice model

Utility for household i in village v from electricity source j in survey wave t is is given by

$$U_{ijtv} = \delta_{jtv} + z'_{it}\gamma_j + \epsilon_{ijt} \quad (5)$$

$$= V_{ijtv} + \epsilon_{ijt} \quad (6)$$

The term V_{ijtv} is the strict utility of a choice for a household absent their idiosyncratic taste shock ϵ_{ijt}. The strict utility depends on the average utility of a source δ_{jtv} as well as a vector z_{it} of observable household characteristics. These characteristics affect household utility through source-specific coefficients γ_j. For example, households with higher incomes may have a greater preference for grid electricity, but an unchanged preference for diesel.

Discrete choice model

The term δ_{jtv} represents the mean utility of an electricity source j in village v at survey wave t. Mean utility depends on observable source characteristics x_{jtv} and unobserved source quality ξ_{jtv}, according to

$$\delta_{jtv} = x'_{jtv}\overline{\beta} + \xi_{jtv} \tag{3}$$

The vector x_{jtv} of observable source characteristics includes price, hours of supply on-peak (from five to ten pm), and hours of supply off-peak. We refer to ξ_{jtv} as unobserved quality or just quality. Unobserved quality is known to households but not the econometrician. It may include

Discrete choice model specification

The authors impose (too?) strong independence assumptons for the ϵ_{ijt}(although allowing for some correlation within nests).

- Different households will weigh price and reliability considerations differently in their energy access choice.
- Choice-specific attributes are calibrated as village-level averages. Household-specific departures from these crude averages are captured in part by the ϵ_{jt}.
- The authors use panel data to estimate the model. But does not allow for state dependence or correlated errors within households across time?

Estimate the model in two stages (recall the Berry two-step)

1. First non-linear stage estimates the mean indirect utilities via maximum likelihood.
2. Second linear stage regresses mean indirect utilities on choice attributes.

The key idea is to invert market shares to solve for mean indirect utilities, which then allows for linear IV estimates in the second stage that are unbiased despite the endogeneity of price to quality (Berry, 1994).

Recall how this works

Solve for the $\beta' X_{jt}$ that fit the observed shares exactly. In the CL model, this inversion step can be done analytically. Taking logs:

$$\log \widehat{s}_0 = -\log(1 + \sum_j \exp(\beta' X_j)) \tag{2}$$

$$\log \widehat{s}_1 = \beta' X_1 - \log(1 + \sum_j \exp(\beta' X_j)) \tag{3}$$

$$\vdots \tag{4}$$

$$\log \widehat{s}_J = \beta' X_J - \log(1 + \sum_j \exp(\beta' X_j)) \tag{5}$$

This yields:

$$\log \widehat{s}_j - \log \widehat{s}_0 = \beta' X_j \tag{6}$$

The Beauty of the Berry transform?

$$\log \widehat{s}_j - \log \widehat{s}_0 = \beta' X_j \tag{7}$$

Intuition: There exists a unique $\beta' X$ such that the predicted shares equal actual (observed) shares.

A key advantage: Inverting market shares to solve for mean indirect utilities allows for linear IV estimates in the second stage that are unbiased despite the endogeneity of price to quality.

First-stage estimation

Non-linear estimation of the first stage. In the first stage, we use maximum likelihood to estimate the parameters δ, γ and σ using equation 4. Let y_{itj} indicate that household i in survey t chose product j. The log-likelihood of the sample is

$$\log \mathcal{L}(\gamma, \sigma | y, z) = \sum_{i=1}^{N} \sum_{t=1}^{T} \log \Pr(y_{itj} | z_{it}; \gamma, \sigma, \delta(\gamma, \sigma)). \tag{5}$$

We write $\delta(\gamma, \sigma)$ to show that we concentrate the δ parameters out of the log-likelihood (Berry, Levinsohn and Pakes, 1995). For every candidate parameter vector (γ, σ) we solve for the δ that exactly fits the aggregate market shares.[25] This greatly reduces the dimensionality of the non-linear search, as the δ vector has up to 1200 elements ($=$ 4 sources $\times$ 100 villages $\times$ 3 surveys), if every source were available in every village.

Second stage regression

In the second stage regression, the mean utility parameters are regressed on price and non-price attributes.

With these instruments, we estimate equation 7 by two stage least squares:

$$\widehat{\delta}_{jtv} = x_{jtv,price}\overline{\beta}_{price} + \sum_{k \neq price} x'_{jtvk}\overline{\beta}_k + \overline{\xi}_{jt} + \tilde{\xi}_{jtv} \quad (11)$$

$$x_{jtv,price} = \pi_1 T_v^{Normal}\mathbf{1}\{Endline\} + \pi_2 T_v^{Subsidized}\mathbf{1}\{Endline\} + \pi_3\widehat{Peak}_{tv} + \pi_4\widehat{OPeak}_{tv} + \overline{\xi}_{jt} + \nu_{jtv}. \quad (12)$$

The authors use the village-level treatment indicators as instruments for price.

Aaron's residual?

Having estimated equation 6, the fitted residuals allow us to recover unobserved quality as

$$\widehat{\xi}_{jtv} = \widehat{\bar{\xi}}_{jt} + \widehat{\tilde{\xi}}_{jtv} = \widehat{\delta}_{jtv} - x'_{jtv}\widehat{\bar{\beta}}.$$

With these estimates, we can observe how the quality of electricity sources varies across sources, villages and time.

Identifying price variation?

The paper is somewhat vague on how monthly prices for othe choices are estimated:

- For grid costs, authors average self-reported monthly payments (problematic)
- For solar PV, investment costs are amortized over 7 years at 20 percent interest rate.

Authors ignore significant variation in these explicit/implicit prices across villages/households.

Thoughts about this IV strategy?

Thoughts about this IV strategy?

- This IV strategy uses only the experimental price variation in a subset of villages to identify the price coefficient.
- Welfare analysis assumes estimated IV price coefficient estimates equally applicable to all energy supply alternatives.
- There are many reasons to think that an incremental change in the microgrid price will be perceived differently as compared to an incremental change in the village monthly average cost of grid connections or PV systems...

Welfare analysis

'With the structural model, we can calculate the surplus from any electricity source by raising the price of that source and calculating the decline in total surplus.'

Table 6: Price Elasticities of Electricity Source Demand

	Elasticity of share for source:				
with respect to price of source:	Grid (1)	Diesel (2)	Own solar (3)	Micro-grid (4)	None (5)
Grid	-0.58	0.31	0.54	0.14	0.14
Diesel	0.06	-1.83	0.24	0.03	0.04
Own solar	0.20	0.44	-1.91	0.12	0.09
Microgrid	0.15	0.16	0.32	-1.58	0.18

The table presents aggregate own- and cross-price elasticities of demand by electricity source. The arc elasticities are calculated using a 10% increase in each source's price from its mean endline price. The elasticities are calculated for the market share of each column source with respect to the price of each row source.

Outline for today

- Welfare measures and the conditional logit model.
- Application 1: Davis (2021)
- Application 2: Burgess et al. (2020)
- Introducing Mixed Logit

Reasons to love the mixed logit framework

1. Extremely general! Researcher can choose the distribution that best fits the empirical setting.
2. Intuitively accommodates random taste variation across people.
3. Accommodates correlation in unobserved attributes over choices/time.
4. No more IIA!

Motivating the mixed logit

Many ways to motivate the mixed logit specification.

The most common (IO, marketing, EEE) and most intuitive (for us economists) starts from the same RUM (or RCM) point of departure.

$$\begin{aligned} U_{ni} &= \beta_n X_{ni} + \varepsilon_{ni} \\ & \varepsilon_{ni}\tilde{\ } iid \text{ EV} \\ & \beta_n\tilde{\ } f(\beta|\theta) \end{aligned}$$

The small but important difference? β can vary randomly across agents. This reflects the fact that different decision-makers have different tastes/preferences.

A small but important difference between ML and CL

- We no longer assume that these taste paramteters are fixed across agents.
- The f(θ) describes the density of these taste coefficientts, where θ is a vector containing the parameters of the distribution of taste parameters.
- For example, if the β is normally distributed, θ contains the mean and variance.
- We now have two types of parameters:
  1. The β_n vary across agents (heterogeneous tastes).
  2. The parameters θ define the assumed distribution in the population.

Conditional choice probabilities

- If we knew the β_n, we could condition on it and get back to the tractable CL choice probability.
- This conditional probability is just the standard logit evaluated at β_n:

$$\mathbb{P}(n \text{ chooses } i \mid \beta_n) = P_{ni}(\beta_n) = \frac{\exp(\beta_n' X_{ni})}{\sum_j \exp(\beta_n' X_{nj})}$$

- This is a nice closed form .. no simulation needed!
- BUT we don't know β_n.. so we can't really condition on it..so what do we do?

Unconditional choice probabilities

To obtain the unconditional choice probability, we need to integrate over the density of β :

$$\mathbb{P}(n \text{ chooses } i) = P_{ni} = \int \frac{\exp(\beta_n' X_{ni})}{\sum_j \exp(\beta_n' X_{nj})} f(\beta) d\beta$$

- This is essentially a weighted average of the logit probabilities evaluated at different β values.
- The weights are given by the density $f(\beta)$.
- This is sometimes called the "mixing" distribution as it defines the weights in this mix of alternative logit functions.
- In most applications, this mixing function is continuous, although there are cases where you might want β to take on a discrete set of values (e.g. latent class models).

Convenient error partitioning

$$
\begin{aligned}
U_{ni} &= \beta_n X_{ni} + \varepsilon_{ni} \\
& \varepsilon_{ni} \tilde{}\ iid \text{ EV} \\
& \beta_n \tilde{}\ f(\beta|\theta)
\end{aligned}
$$

- We continue to assume that the ε are distributed iid EV. So part of the integration can be done analytically.
- This model can be generalized to accommodate both observable and unobservable taste variation.
- An alternative motivation places more emphasis on capturing substitution patterns across choices parsimoniously and realistically.

How do you estimate this thing?

$$
\begin{aligned}
U_{ni} &= \beta_n X_{ni} + \varepsilon_{ni} \\
&\varepsilon_{ni}\tilde{\ } iid \text{ EV} \\
&\beta_n \tilde{\ } f(\beta|\theta)
\end{aligned}
$$

How do you estimate this thing?

$$
\begin{aligned}
U_{ni} &= \beta_n X_{ni} + \varepsilon_{ni} \\
& \varepsilon_{ni}\tilde{}\, iid \text{ EV} \\
& \beta_n \tilde{}\, f(\beta|\theta)
\end{aligned}
$$

- Integrate explicitly over ε_{in} given β_n.
- Integrate numerically (via simulation) over the assumed density $f(\beta_n|\theta)$
- Mechnically, how can we evaluate the integral of a statistic over the assumed density?

Simulation!

If we have a well defined distribution, we can randomly draw from it...

1. Draw R times from the assumed distribution $f(\beta_n|\theta)$.
2. Construct R values of the conditional choice probability evaluated using these R values.
3. Take an average. This gives you your simulated choice probability:

$$P_{ni} = \frac{1}{R}\sum_r \frac{\exp(\beta'_{nr}X_{ni})}{\sum_j \exp(\beta'_{nr}X_{nj})}$$

Insert these simulated probabilities into your simulated log-likelihood function:

$$SLL(\theta) = \sum_n \sum_j y_{nj} \ln\left(\frac{1}{R}\sum_r \frac{\exp(\beta'_{nr}X_{ni})}{\sum_j \exp(\beta'_{nr}X_{nj})}\right)$$

Your ML parameter estimates of θ are those that maximize the value of this simulated likelihood function.

Why is this referred to as convenient error partitioning?

- Partition the error into a component that is distributed in a way that yields choice probabilities that are analytically integrable AND an error that is distributed in a way that makes sense given the choice context.
- Note that an integral over a density is essentially a weighted average.
- We can approximate this with a simulated weighted average.

Distributional assumptions?

- What you assume about the distribution of the random components really depends on the context.
- Suppose you assume the coefficient is distributed randomly in the population : $\beta\tilde{}N(b, s^2)$
- To simulate drawing from this distribution, take a random draw from a standard normal. Multiply by s. Add b. You are done!
- When would you want to choose a distribution *other* than a normal?
- Price coefficient? Log normal a better choice (price enters negatively). Draws from a log normal: $\beta = exp(b + s\eta)$.

Mixed logit to the rescue?

The mixed logit (aka random coefficients logit, error components model) addresses the three limitations we've been discussing.

- Models/accounts for random taste variation/heterogeneity.
- Can accommodate orrelation in unobserved factors across choices/time.
- Substitution patterns uncovered versus imposed.

The key difference: ML model is not tied to any particular distributional assumptions for the unobserved component. Researcher can choose the distribution that best fits the empirical setting.

Panel data

Suppose we observe T_n choice situations/outcomes for person n. Note this number can vary across people. So now we have:

$$\begin{aligned} U_{nti} &= \beta_n X_{nti} + \varepsilon_{nti} \\ & \varepsilon_{nti}\tilde{}\ EV1 \\ & \beta_n \tilde{} f(\beta|\theta) \end{aligned}$$

The outcome of these choice situations is a vector: $y_n = \{y_1 ... y_{T_n}\}$

Conditional on the β for that person, the probability of observing the sequence of decisions is:

$$\Pr ob(y|\beta) = \prod_{t=1}^{T_n} \frac{e^{\beta_n' x_{nty_{nt}}}}{\sum e^{\beta_n' X_{njt}}}$$

Given assumed distribution of ε, this is just the product of logit choice probabilities.

Panel data, unconditional probabilities

The unconditional probability integrates this over the density of β.

$$(y|\theta) = P(y|\beta)f(\beta|\theta)d\beta$$

How does this accommodate serial correlation?

Panel data, unconditional probabilities

The unconditional probability integrates this over the density of β.

$$(y|\theta) = P(y|\beta)f(\beta|\theta)d\beta$$

How does this accommodate serial correlation?

$$\begin{aligned} U_{nti} &= \beta_n X_{nti} + \varepsilon_{nti} \\ &= bX_{nti} + s\eta_n X_{nit} + \varepsilon_{nit} \\ &= bX_{nit} + e_{nit} \end{aligned}$$

- Note that $cov(e_{nit}, e_{nit-1}) = s^2$ depends on the variance of the β and the extent of correlation of the X_{nit} across choices.
- Error component now correlated across choices made by the same agent.

More realistic substution patterns

- The restrictive assumptions of the CL model placed very strong restrictions on substitution patterns.
- Problem stems from the iid errors that are uncorrelated across choices and people.
- With heterogeneous preferences, the agent who chose a more fuel efficient car has a β vector that weights fuel efficiency more heavily than average.
- In other words, error components are correlated across choices with similar choice attributes (Top Dog and LaVals)
- This generates disproportionate substitution beteen more similar choices.

Substitution patterns

Random coefficients/error components that are common across alternatives create correlation in random components across alternatives. To see this, let us reformulate the model just slightly:

$$\begin{aligned} U_{ni} &= \beta_n X_{ni} + \varepsilon_{ni} \\ &= (b + \upsilon_n) X_{ni} + \varepsilon_{ni} \\ &= b X_{ni} + \upsilon_n X_{ni} + \varepsilon_{ni} \\ &= b X_{ni} + e_{ni} \end{aligned}$$

- If you look at the correlation in the errors across alternatives within a decision-maker, they now depend on the choice characteristics.
- The ratio of choice probabilities now depends on all of the data, including attributes of all choice alternatives.

What about those individual-specific parameters?

- Once you have estimated your ML model, you have in hand estimates of the parameters of the distributions of the β coefficients in the population (e.g. WTP for fuel efficiency).
- But you might want to know where a particular agent is within this distribution. (Why??)
- Each agent's choices reveal something about her preferences. We know how her choices differ from others in the population.
- How can we formalize inference from the population to an individual ?

Individual specific parameters

Distinguish between two distributions of the random taste parameters:

1. The distribution of the parameter in the population: $\beta \tilde{} f(\beta|\theta)$. This conditions only on θ.
2. The distribution $h(\beta|y, X, \theta)$ This is the distribution of β in the subpopulation who, when faced with the choice situation characterized by X would make the set of choices y.

The h( ) distribution depends on observed choices y, the choice set X and the population parameters θ.

Loosely speaking (some weighting issues aside), if we sum across all these h distributions, we should get the f (population) distribution back.

Individual taste parameters

$$
\begin{aligned}
U_{nit} &= \beta_n' X_{nit} + \epsilon_{nit} \\
& \varepsilon_{nit}\tilde{}\, iid \text{ EV} \\
& \beta_n\tilde{}\, f(\beta|\theta) inpopulation
\end{aligned}
$$

Outcome vector is y_n. Conditional on persons tastes β_n:

$$P(y_n|\beta) = \prod_{t=1}^{T_n} \frac{e^{\beta_n' X_{nty_{nt}}}}{\sum_j e^{\beta_n' X_{njt}}}$$

Just a product of logit formulas!

The unconditional probability (probability of the choices given θ):

$$P(y_n|\theta) = \int P(y_n|\beta) f(\theta) d\beta$$

Bayes Theorem (not Bayesian estimation!)

What can we learn from agents' choices about the h distribution for each decision maker in our sample (realizing each decision maker belongs to a particular sub-population characterized by y and X).

Bayes rule (which relates conditional and marginal distributions) tells us that the joint probability of β and y can be expressed in two equivalent ways:

1. The probability of β given observed choices y times the probability of observed choices given θ.
2. The probality of the observed choices given β (a conditional distribution) times the probability of β (marginal distribution)

$$h(\beta|y,\theta)P(y|\theta) = P(y|\beta)f(\beta|\theta)$$

This gives us a way to calculate h!

Individual specific parameters

Rearranging this implication of Bayes rule:

$$h(\beta|y_n, X_n, \theta) = \frac{P(y_n|X_n, \beta)f(\beta|\theta)}{P(y_n|X_n, \theta)}$$

- $P(y_n|X_n, \beta)$ is the probability of making the observed choices y given β
- $P(y_n|X_n, \theta)$ is the probability of making the observed choices given the population parameters.
- $f(\beta|\theta)$ is the marginal distribution of β

Recall from the unconditional mixed logit choice probability:

$$P(y_n|X_n, \theta) = \int P(y_n|X_n, \beta)f(\beta|\theta)d\beta$$

So if we make this substitution we have:

$$h(\beta|y_n, X_n, \theta) = \frac{P(y_n|X_n, \beta)f(\beta|\theta)}{\int P(y_n|X_n, \beta)f(\beta|\theta)d\beta}$$

Individual specific parameters

$$h(\beta|y_n, X_n, \theta) = \frac{P(y_n|X_n, \beta)f(\beta|\theta)}{\int P(y_n|X_n, \beta)f(\beta|\theta)d\beta}$$

- Numerator: population density $f(\beta|\theta)$ weighted by the likelihood of a person's observed choices y for each possible value of β.
- Denominator: Integral of numerator (a normalizing constant).
- Intuitively - move through the distribution of β in the population and weight each value by the probability that this agent would have made the choices he made had he had that β.

Individual specific parameters

- This is not Bayesian estimation! This is an implication of your MLE of θ which you can derive/uncover with the help of the identity that is Bayes theorem.
- Cool... but why bother?