

FROM SEQUENTIAL BINARY MODELS TO FIRST-BEST MULTINOMIAL CHOICE: A STATA IMPLEMENTATION

Yunrong Li¹ Ricardo Mora²

2026 UK Stata Conference
3-4 September | London

¹Dali University, China

²Universidad Carlos III Madrid, Spain

Outline

- 1 Complementary binary choices
- 2 From sequential binary choices to bundle choice
- 3 The `compchoice` command: a brief tour on features and usage

Complementary binary choices

Complementary binary choices

- Many economically important binary decisions are **jointly** determined.
 - Studying each decision separately may therefore miss whether the two activities are **complements or substitutes**.
- **Utility complementarity**, γ : undertaking one activity increases the incremental utility from undertaking the other.
 - Gentzkow [2007] implements this idea in a multinomial logit bundle-choice model with i.i.d. Type-I extreme-value shocks.
- We extend this framework in two directions:
 - Singleton-specific shocks are inherited by the joint bundle.
 - Complementarity may contain an unobserved, time-invariant, agent-specific component.
- Our Stata command `compchoice` estimates both bundle-choice specifications and reports, among other quantities, $E(\gamma_{it})$ and $\Pr(\gamma_{it} > 0)$.

Illustration: grandparental childcare and co-residence

- We illustrate `compchoice` using the joint decisions of grandparents to provide childcare and to co-reside with their adult children and grandchildren.
 - Co-residence may facilitate grandparental childcare by reducing distance and coordination costs.
 - Conversely, access to grandparental childcare may increase the value of co-residence for the younger generation.

- There is no natural common hierarchy between the two decisions.
 - Childcare arrangements may shape co-residence in some families, while
 - living arrangements may determine childcare in others.

- With two binary decisions there are four mutually exclusive alternatives:
 - 1 None, $N = (0, 0)$;
 - 2 Only co-Residence, $OR = (1, 0)$;
 - 3 Only childcare-Sharing, $OS = (0, 1)$;
 - 4 Both, $B = (1, 1)$.

From sequential binary choices to bundle choice

Bivariate probit: correlated choices without complementarity

$$D_{\text{co}} = \mathbf{1}\{W_{\text{co}} + v_{\text{co}} > 0\}$$

$$D_{\text{care}} = \mathbf{1}\{W_{\text{care}} + v_{\text{care}} > 0\}$$

with normal correlated shocks v_{co} and v_{care}

- The two decisions are statistically related through common unobserved factors.
- But correlation between the choices is **not utility complementarity**:
 - one activity does not change the utility gain from undertaking the other.
- In fact, when $\gamma = 0$, the bundle-choice multinomial model and the bivariate binary-choice model are equivalent [Gentzkow, 2007].
- Hence, positive choice correlation can arise even when $\gamma = 0$.

Probit with endogenous binary regressor

Suppose co-residence is the endogenous binary explanatory variable:

$$D_{co} = \mathbf{1}\{W_{co} + v_{co} > 0\}$$

$$D_{care} = \mathbf{1}\{W_{care} + \alpha D_{co} + v_{care} > 0\}$$

- α measures how choosing co-residence shifts the latent utility of childcare-sharing.
- This allows one decision to affect the other directly, but imposes a **hierarchy**: co-residence is prior and childcare adjusts later.
 - The relationship is asymmetric: $D_{co} \rightarrow D_{care}$, $D_{care} \nrightarrow D_{co}$.
- By contrast, utility complementarity γ in the bundle-choice model is symmetric and **does not require** an ordering of the two decisions.
- Hence, a sequential representation cannot reproduce all choice patterns allowed by the multinomial bundle model [Li and Mora, 2026].

Two useful results

Gentzkow [2007]

Let z^{co} be an exclusive driver of utility of co-residence. Then

$$\text{sign}\left(\frac{\partial \Pr(D_{care}=1)}{\partial z^{co}}\right) = \text{sign}(\gamma).$$

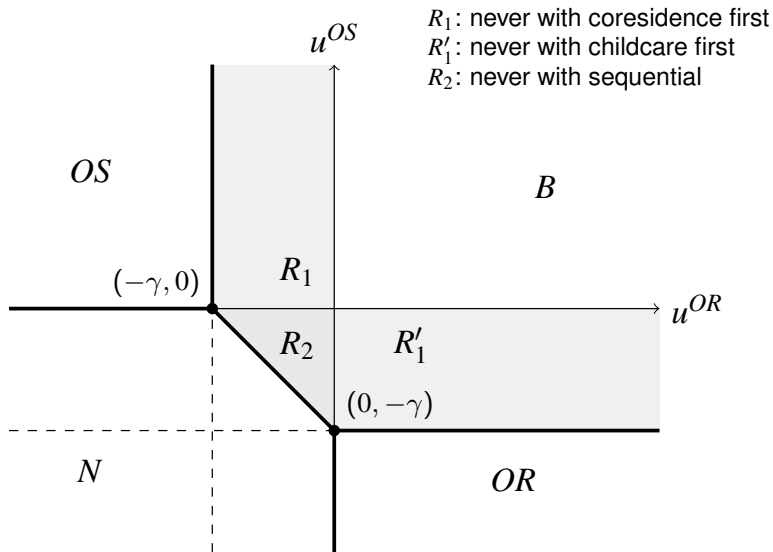
- Utility complementarity is revealed by the **cross effect** of z^{co} on the marginal choice probability of childcare.
- This is the bridge between utility complementarity and Hicksian (demand) complementarity.

Li and Mora [2026]

If $\gamma > 0$, the sequential model and the multinomial bundle-choice model differ only for

- $R_1 = \{(u^{OS}, u^{OR}) \in \mathbb{R}^2 : -\gamma < u^{OR} < 0, u^{OS} > 0\}$ and
- $R_2 = \{(u^{OS}, u^{OR}) \in \mathbb{R}^2 : -\gamma < u^{OR} < 0, -\gamma < u^{OS} < 0, u^{OS} + u^{OR} + \gamma > 0\}$
- With positive complementarity, the co-residence-first sequence fails precisely on $R_1 \cup R_2$
- Similar result for the alternative sequential model.

Choice regions under positive complementarity



The `compchoice` command: a brief tour on features and usage

Basic usage

- The simplest call to `compchoice` specifies the **two binary choices** to be analyzed jointly.
- In our application, the unit of analysis is a grandmother--parent pair observed over multiple waves ($n = 8,793$ pairs, $T \leq 4$). We study the joint decisions to **co-reside** (`coreside`) and to provide **grandparental childcare** (`gpcare`).
- We type

```
. compchoice coreside gpcare, cluster(householdID) ///
      twolevel(householdID)
```

- The two additional options account for the panel structure:
 - `twolevel(householdID)` specifies the level at which the unobserved random effects are shared across observations;
 - `cluster(householdID)` computes standard errors allowing arbitrary dependence within households.
- The order of the two dependent variables defines the singleton alternatives: (1, 0) is co-residence only and (0, 1) is childcare only.

```
compchoice bi_vars, cluster(clst_var) twolevel(tlev_var)
```

First stage: Running Complementarity Multinomial Mixed Logit on estimation sample

Second stage: Performing complementarity analysis.

Complementarity Analysis between coreside and gpcare

Number of obs = 14,046

cores_gpcar	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
E_gamma at_means	.3807352	.0468945	8.12	0.000	.2888237	.4726468
SD_gamma at_means	.5310925	.0785182	6.76	0.000	.3771997	.6849854
Gamma_pos at_means	.746	.0460578	16.20	0.000	.6557283	.8362717

- The benchmark output goes directly to the three main measures of complementarity.
 - Complementarity is positive on average, heterogeneous across families, and positive for roughly three quarters of the estimation sample.
- `compchoice` first estimates the underlying Complementarity Multinomial Mixed Logit model and then performs the simulation-based complementarity analysis.
 - Model estimation coefficients are suppressed by default so that the output focuses on the complementarity measures.
 - The three digit accuracy in the probability is due to the default number of simulations (1,000).

The detail `fulldisplay` options: estimation

First stage: Running Complementarity Multinomial Mixed Logit on estimation sample

Bundle-Choice Multinomial Mixed Logit with Two Latent Factors

Obtaining initial conditions... done.

(setting technique to bhhh)

Iteration 0: $f(p) = -15229.344$

Iteration 1: $f(p) = -14676.003$

Iteration 2: $f(p) = -14584.716$

(switching technique to bfgs)

Iteration 3: $f(p) = -14575.415$

Iteration 4: $f(p) = -14575.291$

Iteration 5: $f(p) = -14575.287$

Iteration 6: $f(p) = -14575.287$

Iteration 7: $f(p) = -14575.287$

Second stage: Performing complementarity analysis.

Bundle-Choice Multinomial Mixed Logit with Two Latent Factors

Number of obs = 14,046

	cores_gpcar	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
2.cores_gpcar							
	_cons	-1.62863	.0453967	-35.88	0.000	-1.717606	-1.539654
3.cores_gpcar							
	_cons	-1.27146	.0401056	-31.70	0.000	-1.350065	-1.192854
4.cores_gpcar							
	_cons	-2.533	.061515	-41.18	0.000	-2.653567	-2.412433
	var(M1[householdID])	1.97128	.157765			1.662066	2.280493
	var(M2[householdID])	2.037006	.1568359			1.729613	2.344399
	cov(M1[householdID],M2[householdID])	.758817	.1102661	6.88	0.000	.5426994	.9749346
	var(epsilon_gamma)	.2419982	.0708632			.1031088	.3808876

Continued on next slide

The detail `fulldisplay` options: cross derivatives

Complementarity Analysis between `coreside` and `gpcare`

Number of obs = 14,046

cores_gpcar		Coefficient	Std. err.	z	P> z	[95% conf. interval]	
E_gamma	at_means	.3807352	.0468945	8.12	0.000	.2888237	.4726468
SD_gamma	at_means	.5310925	.0785182	6.76	0.000	.3771997	.6849854
Gamma_pos	at_means	.746	.0460578	16.20	0.000	.6557283	.8362717
Cross_effect	at_means	.0101559	.0033232	3.06	0.002	.0036426	.0166693
Own_effect_coreside	at_means	.1341925	.008397	15.98	0.000	.1177348	.1506502
Own_effect_gpcare	at_means	.1515579	.0092081	16.46	0.000	.1335103	.1696054
Cross_Own_coreside	at_means	.0756819	.023024	3.29	0.001	.0305556	.1208082
Cross_Own_gpcare	at_means	.0670103	.0219735	3.05	0.002	.0239431	.1100775

Continued on next slide

The detail `fulldisplay` options: family types

coreside_priority_pos at_means	.02	.0056461	3.54	0.000	.0089338	.0310662
coreside_priority_neg at_means	0	.0025656	0.00	1.000	-.0050285	.0050285
gpcare_priority_pos at_means	.02	.0069019	2.90	0.004	.0064726	.0335274
gpcare_priority_neg at_means	.002	.0026563	0.75	0.452	-.0032063	.0072063
Non_sequential at_means	.002	.0032968	0.61	0.544	-.0044615	.0084615
Order_robust_pos at_means	.704	.0468722	15.02	0.000	.6121322	.7958678
Order_robust_neg at_means	.252	.0438537	5.75	0.000	.1660484	.3379516

The `modelonly` option

- With `modelonly` the command **stops after estimation** of the bundle-choice model
- We illustrate this option by adding two covariates, `gfather` and `Dregion4` to the equations.

```
. compchoice coreside gpcare gfather Dregion4, ... modelonly
```

First stage: Running Complementarity Multinomial Mixed Logit on estimation sample

Bundle-Choice Multinomial Mixed Logit with Two Latent Factors

Number of obs = 14,046

cores_gpcar		Coefficient	Std. err.	z	P> z	[95% conf. interval]	
2.cores_gpcar							
	gfather	-.063588	.08796	-0.72	0.470	-.2359865	.1088105
	Dregion4	.4464238	.0827133	5.40	0.000	.2843088	.6085389
	_cons	-1.67228	.0876523	-19.08	0.000	-1.844076	-1.500485
3.cores_gpcar							
	gfather	.6107975	.0912307	6.70	0.000	.4319885	.7896065
	Dregion4	.0910008	.0868012	1.05	0.294	-.0791264	.261128
	_cons	-1.791348	.0901886	-19.86	0.000	-1.968114	-1.614581
4.cores_gpcar							
	gfather	.4783365	.1300758	3.68	0.000	.2233926	.7332804
	Dregion4	.5434091	.1283954	4.23	0.000	.2917587	.7950595
	_cons	-3.041147	.1299128	-23.41	0.000	-3.295772	-2.786523
	var (M1 [householdID])	1.944271	.157194			1.636177	2.252366
	var (M2 [householdID])	1.985146	.1530163			1.685239	2.285052
cov (M1 [householdID],M2 [householdID])		.7616085	.1098318	6.93	0.000	.5463421	.9768748
	var (epsilon_gamma)	.2446029	.0710571			.1053335	.3838724

The modelonly betagama (short) options

First stage: Running Complementarity Multinomial Mixed Logit on estimation sample
 Bundle-Choice Multinomial Mixed Logit with Two Latent Factors

Number of obs = 14,046

cores_gpcar	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
Gamma						
gfather	-.068873	.0799514	-0.86	0.389	-.2255748	.0878288
Dregion4	.0059844	.0716737	0.08	0.933	-.1344935	.1464623
_cons	.422481	.0816411	5.17	0.000	.2624673	.5824947

- `betagama(short)` displays only the implied complementarity equation

$$\beta_{\gamma} = \beta_B - \beta_{OS} - \beta_{OR}$$
- Combined with `modelonly`, it provides a compact view of the **observed sources of complementarity**.
- `betagama(long)` additionally displays the two singleton equations.
- Neither `gfather` nor `Dregion4` appears to be an important observed source of complementarity, even though both affect the underlying choice utilities.

Further options and extensions

- **Evaluation points:** Complementarity measures need not be evaluated only at the default covariate means:
 - `sample` uses the observations in the estimation sample;
 - `conditioning()`, `values()`, and `atvalues()` allow the user to define particular covariate profiles.
- **Model options:** The command also allows alternative assumptions about the bundle-choice model:
 - the Gentzkow specification through `model(gentzkow)`;
 - restrictions on unobserved heterogeneity through `nocovariance`, `equalvariances`, and `novargamma`.
- **Additional features:** These include exclusive singleton drivers through `exc10()` and `exc01()`, general linear restrictions through `constraints()`, and user control over simulation, optimization, and standard table display options.

Thank you!

Matthew Gentzkow. Valuing new goods in a model with complementarity: Online newspapers. *The American Economic Review*, 97(3):713–744, 2007. ISSN 00028282.

Yunrong Li and Ricardo Mora. The complementarity of children and grandparents. Mimeo, 2026.