E673 PS2

Click here for code (Github)

Austin Knies and David Terner

April 19th, 2021

Contents

1 Question 1 (theory: derivation of logit choice probabilities)

Question 1 Setup

Consider the following random utility model. The utility for individual i when choosing alternative j is specified as

$$y_{ij}^* = x_{ij}'\beta + \epsilon_{ij} \; j = 0, \dots, J$$

ϵ_{ij} and x_{ij} are independent. ϵ_{ij} is drawn from a type-1 extreme value distribution. The CDF $F(\epsilon_{ij})$ and the PDF $f(\epsilon_{ij})$ are

$$\begin{aligned} F(\epsilon_{ij}) &= \exp\Big\{-\exp\{-\epsilon_{ij}\}\Big\} \\ f(\epsilon_{ij}) &= \exp\{-\epsilon_{ij}\}\exp\Big\{-\exp\{-\epsilon_{ij}\}\Big\} \end{aligned}$$

We do not observe the utility but only the choice y_i which maximizes i's utility

$$y_i = j \text{ if } y_{ij}^* > y_{ih}^* \; \forall h \neq j.$$

Question: Show that the conditional choice probability $P(y_i = 0|x_i)$ with $x_i = (x_{i0}, \dots, x_{iJ}')$ is given by the logit function:

$$P(y_i = 0|x_i) = \frac{\exp\{x_{i0}'\beta\}}{\sum_{j=0}^{J}\exp\{x_{ij}'\beta\}}$$

Hint: By change of variable,

$$\int_{-\infty}^{\infty} \exp\{-\epsilon\}\exp\Big\{-\exp\{\epsilon - c\}\Big\}d\epsilon = \exp\{c\}$$

holds for any constant c.

1.1 Solution

We follow Ch.3 of Train (2009) for this derivation. We can rewrite the $P(y_i = 0|x_i)$, the probability that individual i chooses alternative $j = 0$, as

$$\begin{aligned} P(y_i = 0|x_i) &= P(y_{i0}^* > y_{ih}^* \;\; \forall h \neq j = 0) \\ &= P(x_{i0}'\beta + \epsilon_{i0} > x_{ih}'\beta + \epsilon_{ih} \;\; \forall h \neq j = 0) \\ &= P(\epsilon_{ih} < \epsilon_{i0} + x_{i0}'\beta - x_{ih}'\beta \;\; \forall h \neq j = 0). \end{aligned}$$

From here, we can clearly see that $P(y_i = 0|x_i)$ is written in terms of the CDF for each ϵ_{ih}, which is equal to $F(\epsilon_{ih}) = \exp\{-\exp\{-\epsilon_{ih}\}\}$. As Train (2009) notes, if ϵ_{i0} were known, and since we assume each ϵ_{ij} to be independent, we could rewrite the probability here in terms of the product of the individual CDFs for each $h \neq j = 0$:

$$\begin{aligned} P(y_i = 0|x_i, \epsilon_{i0}) &= \prod_{h \neq j = 0} F(\epsilon_{ih}) \\ &= \prod_{h \neq j = 0} \exp\Big\{-\exp\Big\{-\Big(\epsilon_{i0} + x_{i0}'\beta - x_{ih}'\beta\Big)\Big\}\Big\} \end{aligned}$$

As Train (2009) points out, ϵ_{i0} is of course *not given*, so instead we must rewrite $P(y_i = 0|x_i)$ as the integral of $P(y_i = 0|x_i, \epsilon_{i0})$ over all values of ϵ_{i0}, weighted by its PDF $f(\epsilon_{i0})$:

$$P(y_i = 0|x_i) = \int_{-\infty}^{\infty} P(y_i = 0|x_i, \epsilon_{i0}) f(\epsilon_{i0}) d\epsilon_{i0}$$

$$P(y_i = 0|x_i) = \int_{-\infty}^{\infty} \left(\prod_{h \neq j = 0} \exp\left\{ - \exp\left\{ - \left(\epsilon_{i0} + x'_{i0}\beta - x'_{ih}\beta \right) \right\} \right\} \right) \times \exp\{-\epsilon_{i0}\} \exp\left\{ - \exp\{-\epsilon_{i0}\} \right\} d\epsilon_{i0}$$

Note that when $h = j = 0$, we have $x'_{i0}\beta - x'_{ih}\beta = 0$. Then we can rewrite the previous equation as

$$\begin{aligned} P(y_i = 0|x_i) &= \int_{-\infty}^{\infty} \left(\prod_{j \in J} \exp\left\{ - \exp\left\{ - \left(\epsilon_{i0} + x'_{i0}\beta - x'_{ij}\beta \right) \right\} \right\} \right) \exp\{-\epsilon_{i0}\} d\epsilon_{i0} \\ &= \int_{-\infty}^{\infty} \exp\left(\sum_{j \in J} - \exp\left\{ - \left(\epsilon_{i0} + x'_{i0}\beta - x'_{ij}\beta \right) \right\} \right) \exp\{-\epsilon_{i0}\} d\epsilon_{i0} \\ &= \int_{-\infty}^{\infty} \exp\left(- \exp\left\{ - \epsilon_{i0} \right\} \sum_{j \in J} \exp\left\{ - \left(x'_{i0}\beta - x'_{ij}\beta \right) \right\} \right) \exp\{-\epsilon_{i0}\} d\epsilon_{i0}. \end{aligned}$$

Define the change of variable $t = \exp\{-\epsilon_{i0}\}$ such that $dt = -\exp\{-\epsilon_{i0}\}d\epsilon_{i0}$. As ϵ_{i0} approaches $-\infty$, t approaches ∞; on the other hand, as ϵ_{i0} approaches ∞, t approaches 0. We rewrite the previous expression using t and the redefined limits:

$$\begin{aligned} P(y_i = 0|x_i) &= \int_{\infty}^{0} \exp\left(- t \sum_{j \in J} \exp\left\{ - \left(x'_{i0}\beta - x'_{ij}\beta \right) \right\} \right) (-dt) \\ &= \int_{0}^{\infty} \exp\left(- t \sum_{j \in J} \exp\left\{ - \left(x'_{i0}\beta - x'_{ij}\beta \right) \right\} \right) (dt) \\ &= \left. \frac{\exp\left(- t \sum_{j \in J} \exp\left\{ - \left(x'_{i0}\beta - x'_{ij}\beta \right) \right\} \right)}{- \sum_{j \in J} \exp\left\{ - \left(x'_{i0}\beta - x'_{ij}\beta \right) \right\}} \right|_{0}^{\infty} \\ &= 0 - \frac{1}{- \sum_{j \in J} \exp\left\{ - \left(x'_{i0}\beta - x'_{ij}\beta \right) \right\}} \\ &= \frac{\exp\{x'_{i0}\beta\}}{\sum_{j=0}^{J} \exp\{x'_{ij}\beta\}} \end{aligned}$$

as we wanted to show. ■

2 Question 2 (theory)

Question 2 Setup

A latent variable y_i^* is generated by

$$\begin{aligned} y_i^* =& x_i\beta + \epsilon_i \\ \epsilon_i | x_i \sim& \mathcal{N}(0, \sigma_i^2) \\ \sigma_i^2 =& \exp\{\gamma x_i\} \end{aligned}$$

where x_i is a scalar. We observe

$$y_i = \begin{cases} 1 & \text{if } y_i^* \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Questions:

(a) Write down the log-likelihood function for the conditional distribution of y_i given x_i.

(b) Find the asymptotic distribution of the maximum likelihood estimator.

(c) Explain how to evaluate the effect of x_i on the choice probability $P(y_i = 1|x_i)$.

(d) Explain how to check heteroskedasticity in the error term, i.e., test that $\gamma = 0$.

2.1 Solution

(a) The latent error term can be re-expressed as:

$$\epsilon \sim \mathcal{N}(0, \exp\{\gamma x_i\})$$

This leads us to a "nice"-looking log-likelihood function:

$$\mathcal{L}(\beta, \gamma) = \sum_{i=1}^{n} \left[y_i \log \left(\Phi \Big(\frac{x_i\beta}{\sqrt{\exp\{x_i\gamma\}}} \Big) \right) + (1 - y_i) \log \left(1 - \Phi \Big(\frac{x_i\beta}{\sqrt{\exp\{x_i\gamma\}}} \Big) \right) \right] \tag{1}$$

Equation (1) is similar to the more conventional equation (4) with one key difference: re-scaling of the RHS of the original latent variable equation. For the time being, assume that $\gamma = 0$ and consider the following "standard" probit derivation from the latent model:

$$\begin{aligned} \Pr(y_i = 1|x) =& \Pr(y_i^* > 0) \\ =& \Pr(x_i\beta + \epsilon_i \geq 0) \\ =& \Pr(\epsilon_i \geq -x_i\beta) \\ =& \Pr(\epsilon_i < x_i\beta) \\ =& \Phi(x\beta) \end{aligned}$$

where the penultimate step (in red) depends on the symmetry of the normal distribution. By scaling all terms by $\sqrt{\exp\{x_i\gamma\}}$ gets our error term back to a **standard normal process** (which allows us to make the penultimate step).

(b) Assuming the standard regularity conditions hold, we know the score of the log likelihood for observation i is

$$\mathbf{s}_i(\beta) = \frac{\partial \mathcal{L}(\beta,\gamma)}{\partial \beta} = y_i\left(\frac{\phi\Big(\frac{x_i\beta}{\sqrt{\exp\{x_i\gamma\}}}\Big)\frac{x_i}{\sqrt{\exp\{x_i\gamma\}}}}{\Phi\Big(\frac{x_i\beta}{\sqrt{\exp\{x_i\gamma\}}}\Big)}\right) - (1-y_i)\left(\frac{\phi\Big(\frac{x_i\beta}{\sqrt{\exp\{x_i\gamma\}}}\Big)\frac{x_i}{\sqrt{\exp\{x_i\gamma\}}}}{1-\Phi\Big(\frac{x_i\beta}{\sqrt{\exp\{x_i\gamma\}}}\Big)}\right)$$
$$= \frac{\phi\Big(\frac{x_i\beta}{\sqrt{\exp\{x_i\gamma\}}}\Big)x_i u_i}{\Big[\sqrt{\exp\{x_i\gamma\}}\Big]\Phi\Big(\frac{x_i\beta}{\sqrt{\exp\{x_i\gamma\}}}\Big)\Big[1-\Phi\Big(\frac{x_i\beta}{\sqrt{\exp\{x_i\gamma\}}}\Big)\Big]}$$

where $u_i = y_i - \Phi(\frac{x_i\beta}{\sqrt{\exp\{x_i\gamma\}}})$. Note that

$$\beta_0 = \arg\max_{\beta\in\Theta} \mathcal{L}(\beta,\gamma)$$

and, thus, under the assumed regularity conditions, we have

$$\mathbf{s}_i(\beta_0) = \frac{\phi\Big(\frac{x_i\beta_0}{\sqrt{\exp\{x_i\gamma\}}}\Big)x_i u_i}{\Big[\sqrt{\exp\{x_i\gamma\}}\Big]\Phi\Big(\frac{x_i\beta_0}{\sqrt{\exp\{x_i\gamma\}}}\Big)\Big[1-\Phi\Big(\frac{x_i\beta_0}{\sqrt{\exp\{x_i\gamma\}}}\Big)\Big]}$$

where $u_i = y_i - \Phi(\frac{x_i\beta_0}{\sqrt{\exp\{x_i\gamma\}}}) = y_i - \mathbb{E}[y_i|x_i]$. Since $\mathbb{E}[u_i|x_i] = 0$, we see that $\mathbb{E}[\mathbf{s}_i(\beta_0)|x_i] = 0$. Following Chapter 13 of Wooldridge (2010), we have that

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}(0, \mathbf{A}^{-1}) \tag{2}$$

where $\mathbf{A}(x_i,\beta_0) = -\mathbb{E}[\mathbf{H}(y_i,x_i,\beta_0)|x_i]$, where $\mathbf{H}_i(\beta)$ is the Hessian matrix of $\mathcal{L}(\beta,\gamma)$ for observation i. At β_0, we have

$$\mathbf{A} = -\mathbb{E}[\mathbf{H}(y_i,x_i,\beta_0)|x_i] = \frac{\Big[\phi\Big(\frac{x_i\beta_0}{\sqrt{\exp\{x_i\gamma\}}}\Big)\Big]^2 x_i^2 e^{-x_i\gamma}}{\Phi\Big(\frac{x_i\beta_0}{\sqrt{\exp\{x_i\gamma\}}}\Big)\Big[1-\Phi\Big(\frac{x_i\beta_0}{\sqrt{\exp\{x_i\gamma\}}}\Big)\Big]}$$

since x_i is scalar.

(c) Based on our scaling, we can write the desired probability as:

$$\Pr(y_i = 1|x_i) = \Phi\Big(\frac{x_i\beta}{\sqrt{\exp\{x_i\gamma\}}}\Big)$$

Taking a partial derivative with respect to x_i leads to:

$$\begin{aligned}\frac{\partial \Pr(y_i = 1|x_i)}{\partial x_i} =& \phi\Big(\frac{x_i\beta}{\sqrt{\exp\{x_i\gamma\}}}\Big)\beta\Big(\frac{1}{(\exp\{x_i\gamma\})^{\frac{1}{2}}} - \frac{x_i\gamma\exp\{x_i\gamma\}}{2(\exp x_i\gamma\})^{3/2}}\Big)\\ =& \beta\Big(\frac{2-x_i\gamma}{2\sqrt{\exp\{x_i\gamma\}}}\Big)\phi\Big(\frac{x_i\beta}{\sqrt{\exp\{x_i\gamma\}}}\Big)\end{aligned}$$

Since $\phi\Big(\frac{x_i\beta}{\sqrt{\exp\{x_i\gamma\}}}\Big) \geq 0$ and $2\sqrt{\exp\{x_i\gamma\}} > 0$, then $\frac{\partial \Pr(y_i=1|x_i)}{\partial x_i} > 0$ if:

$$\beta\Big(\frac{2 - x_i\gamma}{2\sqrt{\exp\{x_i\gamma\}}}\Big) > 0 \implies \beta(2 - x_i\gamma) > 0$$

If we assume that $\gamma > 0$, then $\frac{\partial \Pr(y_i=1|x_i)}{\partial x_i} > 0$ iff:

Case (a) $\beta > 0$

(i) $x < 0$

(ii) $0 < x < \frac{2}{\gamma}$

Case (b) $\beta < 0$

(i) $x > 0$

(ii) $0 > x > -\frac{2}{\gamma}$

otherwise an increase in x_i will decrease the probability of the latent variable realization. At any rate, assessing x_i's effect on choice probability is far more complicated than a more vanilla probit model. Not only does $x_i's$ marginal impact hinge on β, but also depends on the sign of x_i and magnitude of the non-constant variance scaling factor γ.

(d) Following the example of Wooldridge (2010), testing for heteroskedasticity in the error term, namely $\gamma = 0$, is straightforward. All we need do is run a simple LR test, with the following test statistic:

$$LR = 2 \times (ll^u - ll^r) \sim \chi^2(p)$$

where ll^u is the log-likelihood score from Equation (1) (e.g. *unrestricted* model) and ll^r is the loglikelihood score from the baseline/*restricted* model.[1] The LR test statistic is chi-square distributed on p degrees of freedom, where p is the number of parameters that are omitted when going from the unrestricted to restricted model (in context $p = 1$).

3 Question 3 (theory: normalization in DCM

Question 3 Setup

Consider a general discrete choice model with $J = 4$ alternatives. Assume an economic model specifies that the covariance matrix of the error term $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{i4})'$ takes the following form

$$\Omega_i = \begin{pmatrix} 1+\rho_1 & \rho_1 & 0 & 0 \\ \rho_1 & 1+\rho_1 & 0 & 0 \\ 0 & 0 & 1+\rho_2 & \rho_2 \\ 0 & 0 & \rho_2 & 1+\rho_2 \end{pmatrix}$$

Question: Are the parameters ρ_1 and ρ_2 identified? Derive your answer using the method discussed in class and explain.
Note: *We may not have time to discuss this method explicitly in class. You can either derive it from our general discussion of scale normalization in discrete choice models or read about it in the probit chapter of the Train textbook.*

[1] In context, this would mean the log-likihood score from a simple probit regression model with constant variance.

3.1 Solution

Following Train (2009) and our class notes, we can use Ω_i to normalize the probit model and determine what parameters are identified. Like Train, we can take differences with respect to the first alternative and find $\tilde{\Omega}_1 = M_1 \Omega_i M_1'$:

$$
\begin{aligned}
\tilde{\Omega}_1 &= \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1+\rho_1 & \rho_1 & 0 & 0 \\ \rho_1 & 1+\rho_1 & 0 & 0 \\ 0 & 0 & 1+\rho_2 & \rho_2 \\ 0 & 0 & \rho_2 & 1+\rho_2 \end{pmatrix} \begin{pmatrix} -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1-\rho_1 & -\rho_1 & 1+\rho_2 & \rho_2 \\ -1-\rho_1 & -\rho_1 & \rho_2 & 1+\rho_2 \end{pmatrix} \begin{pmatrix} -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2+\rho_1+\rho_2 & 1+\rho_1+\rho_2 \\ 1 & 1+\rho_1+\rho_2 & 2+\rho_1+\rho_2 \end{pmatrix}
\end{aligned}
$$

To scale utility, we normalize the top-left element of $\tilde{\Omega}_1$, which happens to be the variance of error difference of alternative $j = 2$ with alternative $j = 1$. In doing so, we obtain

$$
\tilde{\Omega}_1^* = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{2+\rho_1+\rho_2}{2} & \frac{1+\rho_1+\rho_2}{2} \\ \frac{1}{2} & \frac{1+\rho_1+\rho_2}{2} & \frac{2+\rho_1+\rho_2}{2} \end{pmatrix}
$$

Here, we note that for parameters ρ_1 and ρ_2 to be identified, it must be possible to estimate each. Now, since there are 4 elements in $\tilde{\Omega}_1^*$ to be estimated, but 2 of them are identical to each other. And given that there are only 2 unique parameters entering Ω_i, ρ_1 and ρ_2, to be estimated, we might think we have a system of two equations and two unknowns. However, as Train points out, it is impossible to estimate ρ_1 and ρ_2 separately. To see this more clearly, we rewrite

$$
\tilde{\Omega}_1^* = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \theta & \theta - \frac{1}{2} \\ \frac{1}{2} & \theta - \frac{1}{2} & \theta \end{pmatrix}
$$

where $\theta = \frac{2+\rho_1+\rho_2}{2}$. Here, it is clear to see that ρ_1 and ρ_2 only appear in the normalized model as the average of the two, and so the average (equal to $\theta - 1$) can be identified. However, ρ_1 and ρ_2 cannot be identified.

4 Question 4 (empirical: binary logit and probit)

Question 4 Setup

This exercise uses `swiss.csv` which contains the following data for 872 Swiss women.

- Column 2: binary labor force participation variable (y_i)
- Column 3: log of non-labor income (I_i)
- Column 4: age in decades (A_i)
- Column 5: education in years (E_i)
- Column 6: number of children under 7 years of age (nu_i)
- Column 7: number of children over 7 years of age (no_i)
- Column 8: citizenship dummy variable (1 if not Swiss) (F_i)

Note: *Given that we will probably not have time to talk about probit models explicitly, parts (d) and (e) are for extra credit only.*
Questions:

(a) Consider a model with y_i as the dependent variable. The regressors are all variables in the data plus A_i^2. Based on this specification, estimate the linear probability model, i.e. the OLS regression of $y_i = x_i'\beta + \epsilon_i$.

(b) Based on the same specification as in (a), estimate the corresponding logit model. Compare the estimates with those of (a). Try not to use ready-to-use package such as Stata's `logit` command but try to code the model yourself in a language like MATLAB or Python. You do not have to compute standard errors. Afterwards, use Stata of R to check your results.

(c) Using the estimate of the logit model in (b), evaluate the marginal effect of education E_i on labor force participation.

(d) Based on the same specification as in (a), estimate the corresponding probit model model assuming that the error terms are *iid* $\mathcal{N}(0, 1)$. Compare the estimates with those of (a) and (b). Try not to use a ready-to-use package such as Stata's `probit` command but code the model yourself in a language like MATLAB or Python. You do no have to compute standard errors. If you get stuck at some point, explain what goes wrong with your code and use Stata or R to estimate the probit model.

(e) Using the estimates of the probit model in (d), evaluate the marginal effect of education E_i on labor force participation.

4.1 Solution

4.1.1 Logistic MLE

Define response variable Y_i as an independent random Bernoulli variable parameterized by ρ_i. The corresponding log-odds can be written as:

$$\log \frac{\rho_i}{1-\rho_i} = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$$

where we have $p+1$ predictors (or p covariates). From this we can re-write the log-odds ratio as:

$$\Pr[Y_i = 1] = \rho_i = \frac{\exp\{\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}\}}{1 + \exp\{\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}\}}$$

Given our definition of ρ_i, the corresponding likelihood function for the logistic regression model is given by:

$$L(\beta) = \prod_{i=1}^{n} \rho_i^{Y_i}(1-\rho_i)^{1-Y_i} = \prod_{i=1}^{n}(1-\rho_i)\Big(\frac{\rho_i}{1-\rho_i}\Big)^{Y_i}$$

Apply a log-transform and rearranging terms yields our desired log-likelihood function:

$$l(\beta)^{\text{logit}} = \sum_{i=1}^{n}\Big[- y_i \log\Big(1 + \exp\{-\sum_{j=0}^{p-1} x_{ij}\beta_j\}\Big) - (1-y_i)\log\Big(1+\exp\{\sum_{j=0}^{p-1} x_{ij}\beta_j\}\Big)\Big] \tag{3}$$

4.1.2 Probit MLE

By a similar sequence of steps, one can show that the Probit MLE is:

$$l(\beta)^{\text{probit}} = \sum_{i=1}^{n}\left[y_i \log\left(\Phi\Big(\sum_{j=0}^{p-1} x_{ij}\beta_j\Big)\right) + (1-y_i)\log\left(1 - \Phi\Big(\sum_{j=0}^{p-1} x_{ij}\beta_j\Big)\right)\right] \tag{4}$$

4.1.3 Estimation Details

We estimate equations (3) and (4) using R's `nlm` and `optim`. We check our coding results against Base R's logit/probit models using the `glm` function. `glm` uses Fisher Scoring as its optimization method. In practice, Fisher Scoring and Newton-Raphson type optimization methods are quite similar. `nlm` uses a Newton-Raphson whereas `optim` implements a Nelder-Mead simplex based method. Based on combination of performance measures—namely AIC, BIC, and McFadden's pseduo-R^2—`nlm` seems to do better than `optim`.[2]

[2]Note: McFadden's pseduo-R^2 is calculated as 1 less the ratio of the log-likelihood score of the fitted model to the log-likihood score of a null model (e.g. a model with only the intercept term).

4.1.4 Linear Probability Model

	Model 1
(Intercept)	1.664^{***}
	(0.446)
income	-0.213^{***}
	(0.041)
age	0.683^{***}
	(0.130)
I(age^2)	-0.097^{***}
	(0.016)
education	0.007
	(0.006)
children_under	-0.241^{***}
	(0.031)
children_over	-0.049^{***}
	(0.017)
as.factor(citizen)1	0.250^{***}
	(0.040)
R^2	0.193
Adj. R^2	0.186
Num. obs.	872
RMSE	0.450

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 1: Linear Probability Model

4.1.5 Logit Regression Model

	Base R (`glm`)	Own Model (`nlm`)	Own Model (`optim`)
(Intercept)	6.196***	6.199***	4.452*
	(2.383)	(2.383)	(2.450)
age	3.437***	3.437***	1.292*
	(0.688)	(0.688)	(0.724)
I(age^2)	−0.488***	−0.488***	−0.211**
	(0.085)	(0.085)	(0.088)
income	−1.104***	−1.104***	−0.766***
	(0.226)	(0.226)	(0.224)
education	0.033	0.033	0.188***
	(0.030)	(0.030)	(0.035)
children_under	−1.186***	−1.186***	−1.492***
	(0.172)	(0.172)	(0.200)
children_over	−0.241***	−0.241***	0.004
	(0.084)	(0.084)	(0.088)
as.factor(citizen)1	1.168***	1.168***	3.475***
	(0.204)	(0.204)	(0.300)
AIC	1033.570	1033.570	1152.789
BIC	1071.736	1071.736	1190.956
Log Likelihood	-508.785	-508.785	-568.395
Deviance	1017.570	1017.570	1136.789
Num. obs.	872	872	872
McFadden pseudo-R^2		0.154	0.055

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 2: Logit Model Comparison

4.1.6 Logit Marginal Effects

Predictor	Average ME	Std Error	z-values	p-values	lower	upper
age	-0.082	0.017	-4.910	0.000	-0.115	-0.049
children_over	-0.048	0.017	-2.902	0.004	-0.081	-0.016
children_under	-0.237	0.031	-7.672	0.000	-0.297	-0.176
citizen1	0.243	0.041	5.977	0.000	0.163	0.323
education	0.007	0.006	1.092	0.275	-0.005	0.018
income	-0.220	0.043	-5.143	0.000	-0.304	-0.136

Table 3: Logit Average Marginal Effects (from `Margins`)

4.1.7 Probit Regression Model

	Base R (`glm`)	Own Model (`nlm`)	Own Model (`optim`)
(Intercept)	3.749^{***}	3.751^{*}	-0.347
	(1.407)	(2.153)	(2.012)
age	2.075^{***}	2.076^{***}	-0.109
	(0.405)	(0.622)	(0.582)
I(age^2)	-0.294^{***}	-0.294^{***}	-0.010
	(0.050)	(0.076)	(0.071)
income	-0.667^{***}	-0.667^{***}	0.124
	(0.132)	(0.200)	(0.183)
education	0.019	0.019	-0.036
	(0.018)	(0.028)	(0.027)
children_under	-0.714^{***}	-0.714^{***}	-0.307^{**}
	(0.100)	(0.153)	(0.143)
children_over	-0.147^{***}	-0.147^{*}	0.002
	(0.051)	(0.079)	(0.076)
as.factor(citizen)1	0.714^{***}	0.714^{***}	-0.232
	(0.121)	(0.188)	(0.181)
AIC	1033.155	1033.155	1203.297
BIC	1071.321	1071.321	1241.463
Log Likelihood	-508.577	-508.577	-593.648
Deviance	1017.155	1017.155	1187.297
Num. obs.	872	872	872
McFadden-psuedo-R^2		0.155	0.013

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4: Probit Model Comparison

4.1.8 Probit Marginal Effects

Predictor	Average ME	Std Error	z-values	p-values	lower	upper
age	-0.083	0.017	-4.965	0.000	-0.116	-0.050
children_over	-0.049	0.017	-2.929	0.003	-0.081	-0.016
children_under	-0.237	0.030	-7.791	0.000	-0.296	-0.177
citizen1	0.246	0.040	6.081	0.000	0.167	0.325
education	0.006	0.006	1.073	0.283	-0.005	0.018
income	-0.221	0.042	-5.275	0.000	-0.303	-0.139

Table 5: Probit Average Marginal Effects (from `Margins`)

5 Question 5 (empirical: demand estimation a la BLP)

Question 5 Setup

This question use the dataset `automobile.csv` containing the data on the automobile industry in Germany from 1991 to 1998. The data include the car model-level information on the number of units sold in each year and each automobile's characteristics.

Questions

1. **Descriptive Statistics**
   Construct a table that shows the number of models, total quantity sold, quantity-weighted average price, and average price characteristics by year, similarly to Table 1 in Berry, Levinsohn and Pakes (1995). You do not have to list all characteristics, although you should include the variables that you use in the following estimations.

2. **Logit Model**
   Assume that the utility function is given by

$$u_{ijt} = X_{jt}\beta - \alpha p_{jt} + \xi_{jt} + \epsilon_{ijt}$$

   where X_{jt} is a vector of observed characteristics, ξ_{jt} is an unobserved product characteristics, p_{jt} is automobile j's price, and ϵ_{ijt} is an *iid* utility shock. Here X_{jt} should include a constant and at least two more variable the choice of which is up to you. Suppose you estimate the model using only the demand side similar to the baseline model in Berry (1994).

   (a) *I suggest you do this part in Stata, as it is convenient and useful to check your MATLAB code later.*
      i. Define the mean utility and compute it for each product.
      ii. Estimate a logit model by regressing the mean utility on the characteristics and the price.
      iii. Create the within category share using the variable `cla`.
      iv. Estimate a nested logit model by regressing the mean utility on the characteristics, price, and the within category share.

   (b) i. Explain the intuition behind BLP-instruments and define the formally.
      ii. Estimate the same logit model as in (a) but use an IV regression with some BLP instruments. The specifics of the instruments are up to you.
      iii. Using the estimated parameters from (a) and (b), calculate own- and cross-price elasticities. Pick any 10 automobiles from 1995, and show price elasticities (in a 10×10 matrix) for model (a) and model (b).

   (c) i. Assume these data were generated under Bertrand-Nash price competition. Each model is produced by the firm reported in the variable `frm`. Compute the price-cost margins for automobiles in 1995.
      ii. Write a table listing the 5 best and 5 worst automobiles in terms of the price-cost margin. Also include their prices and some characteristics.

3. **Random coeffcients logit model**
   Now assume that the utility function is given by

$$u_{ijt} = X_{ijt}\beta_i = \alpha_i p_{jt} + \xi_{jt} + \epsilon_{ijt}$$

where $\beta_{ij} = \bar{\beta}_k + \beta_k^u \nu_{ij}$ and $\nu_{ij} \sim \mathcal{N}(0,1)$. Use the same variables as in the previous question. Suppose two variables (the choice is up to you) have random coefficients. The other variables do no have random coefficients. *I strongly suggest you do this part in MATLAB, as it is very tedious to program this part in Stata.*

(a) i. Estimate the parameters, $(\alpha, \bar{\beta}, \beta_1^u, \beta_2^u)$, using the BLP-method. As a weighting matrix use the standard 2SLS matrix $(Z'Z)^{-1}$ where Z is your matrix of instruments. *I suggest that use* `fminsearch` *in MATLAB to minimize the GMM objective function (once you have an idea on the magnitude of the coefficients, you might want to try* `fminunc` *for a more efficient gradient-based optimization).*

ii. Report your point estimates and the GMM objective value. You do not have to compute standard errors. If you the "Nevo-trick" from Nevo (2000) of separating linear and nonlinear parameters:

A. Have brand dummies in your product characteristics; and,

B. Use a zero vector as your starting value for the linear parameters.

(b) i. Use the estimates, calculate own- and cross-price elasticities.

ii. Report price elasticities (10×10 matrix) for 10 automobiles in 1995.

(c) Assume the data were generated under Nash-Betrand price competition. Each model is produced is produced by `frm` in the data set.

i. Back out the marginal cost for the automobiles i 1995.

ii. Construct a table listing the 5 best and the 5 worst automobiles in terms of the price-cost margin. Also report prices and some product characteristics.

4. **Comparison**
Suppose you successfully estimate two models above (?!?!).

(a) Compare your (a) estimates, (b) own- and cross-price elasticities and (c) price-cost margins from the logit model in Question 2 and the random coefficients logit model in Question 3.

(b) Discuss the differences, e.g., whether or not these results are reasonable. If you fail to estimate the models, you can discuss hypothetical differences based on your expected results.

5. **Income Effect**.
Now assume that the utility function is given by

$$u_{ijt} = X_{jt}\beta_i - \alpha \log(y_{it} - p_{jt}) + \xi_{jt} + \epsilon_{ijt}$$

where y_{it} denotes the income of household i in year t. Assume that the mean income was 20,000 EUR and the standard deviation is 20,000 EUR. Briefly explain how to estimate the model. In your explanation you need to

(a) Clarify how to use the information about the income distribution; and,

(b) Define elements of the substitution matrix, Δ. You do **not** have to estimate this model.

6. **Merger Analysis**
Assume the data were generated under Nash-Bertrand price competition. Suppose all German companies, i.e., BMW, Mercedes, and VW, merge and become one profit-maximizing company.

(a) How would you compute the new pricing equilibrium?
(b) Describe how to compute it under the logit demand specification.
(c) Repeat this exercise for the:
   i. random coefficients; and the
   ii. logit demand specification.

   You do **not** have to compute the new equilibrium.

5.1 Solution

5.1.1 Descriptive Statistics

Year	Number of Models	Sales	Average Price	HP/Wt	Size	Speed
91	76	2758.46	25.81	0.06	7.04	168.11
92	75	3089.40	24.40	0.05	6.96	165.70
93	74	2724.97	27.07	0.05	7.05	168.18
94	86	2802.30	27.05	0.05	7.04	168.32
95	86	2895.20	27.78	0.05	7.05	168.54
96	90	2912.82	28.74	0.05	7.06	169.79
97	89	3035.54	30.34	0.05	7.14	172.83
98	92	3111.93	30.88	0.06	7.19	174.45

Table 6: BLP Table I Analog
Notes: *Number of models* refers to the distinct number of `co` instances per year. *Sales* (`qu`) refers to the number of new car registrations per year (in '000s). The remaining columns of sales (`qu`) weighted averages. *Average Price* measures in destination currency '000s, *HP/Wt* measures the average horse power to weight ratio, *Size* measures the average size (length times width) in square meters, and *Max Speed* measures the average max speed (km/hr).

Quartile	Price	Sales	HP/Wt	Size	Speed
Min	fiat panda	lancia dedra	seat ibiza	lancia dedra	fiat panda
	9.99	0.378	0.035	4.301	125
Q25	mazda 323	renault 5	volkswagen passat	renault 5	citroen AX
	19.16	3.61575	0.05	6.303	158
Median	renault 21	seat marbella	nissan sunny	seat marbella	volvo 940-960
	26.208	13.024	0.059	7.241	173
Q75	hyundai sonata	mazda 626	mercedes 190	mazda 626	ford scorpio
	36.4175	37.115	0.069	7.968	192
Max	audi a8	volkswagen golf	honda civic	volkswagen golf	audi a8
	99.7	414.132	0.11	9.466	247

Table 7: BLP Table II Analog
Notes: All variables have the same definition as in Table 6. The model closest to each variable's listed quartile is listed above the quartile value.

5.1.2 Logit Model

(a) Mean utility of car j at time t is defined as:

$$\delta_{jt} = X_{jt}\beta - \alpha p_{jt} + \xi_{jt} \tag{5}$$

where design matrix X_{jt} includes a column of one's, a ratio of horse power to weight (e.g. HP/Wt), the size of the car (e.g *Size*), and the max speed of the car (e.g. *Speed*). Appealing to the class lecture notes and Berry (1994), we know that mean utility δ_{jt} can easily be recovered from data. Define market size s_{jt} as $s_{jt} = q_{jt}/M_t$, where q_{jt} is the total quantity of model j sold at time t: this approach uses variable `qu` for q_{jt}. Likewise, define $M_t = \sum_{j=0}^{J} q_{j,t}$. Finally, define outside market share as $s_{0t} = 1 - \sum_{j=1}^{J}$. With these definitions in mind, re-define δ_{jt} as:

$$\delta_{jt} = \ln(s_{jt}) - \ln(s_{0t}) \tag{6}$$

Finally, we use OLS to estimate:

$$\delta_{jt} = \beta_0 + \beta_1 hp\_wt_{ijt} + \beta_2 size_{ijt} + sp_{ijt} + \alpha p_{ijt} + \varepsilon_{ijt} \tag{7}$$

For the next part of this question question, we define a within category share as:

$$s_{jtc} = \sum_{c \in C} \frac{q_{jtc}}{M_{tc}}$$

where c is a grouping variable corresponding to all instances of `cla` (e.g. class or segment code). In words, s_{jtc} is j's share of units sold within class c at time t. We update equation (7) to include s_{jtc} as a RHS predictor.

The results for both regression models are reported below in Table 8. Please note that $s_{jtc} \in (0, 1)$ and that *Price* has been scaled down by a factor of $1,000$.

Table 8: Mean Utility Regression from Equation (7)

	Dependent variable:		
	δ_{jt}		
	(Baseline)	(Nested 1)	(Nested 2)
hp_wt	-65.234^{***}	-23.778^{***}	-3.630^{*}
	(8.012)	(5.517)	(2.160)
size	0.363^{***}	0.433^{***}	0.076^{**}
	(0.117)	(0.078)	(0.030)
sp	-0.014^{*}	-0.014^{***}	-0.002
	(0.008)	(0.005)	(0.002)
pr_s	0.017^{*}	-0.031^{***}	-0.027^{***}
	(0.010)	(0.007)	(0.003)
s_{jtc}		11.179^{***}	
		(0.386)	
$\log(s_{jtc})$			0.989^{***}
			(0.010)
Constant	-2.179^{***}	-4.344^{***}	-0.884^{***}
	(0.710)	(0.478)	(0.183)
Observations	668	668	668
R^2	0.247	0.667	0.950
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Notice that the price coefficient α for the baseline model is both *positive* and *significant* at the 10% significance level. This would suggest upward sloping demand curves...which perhaps we can attribute to missing quality variation. By contrast, once we control for within category/class share, the price coefficient α is both *negative* and *significant* at the 1% significance level. The last column uses $\log(s_{jtc})$ as a predictor to follow along with the lecture slides. (Nested 2) results are much the same as (Nested 1) results, especially insofar as the sign and significance of the price coefficient are concerned.

(b) The intuition behind BLP instruments is to find a set of plausible stand-ins for cost shifters that can be used to eliminate price endogeneity. Moreover, it must be the case that: (i) conditioning on this set of instruments, the expected (unobserved) quality dimension ξ_j must be zero; and, (ii) the set of instruments must not correlate with the price of j. For this example, we define our set of instruments to be the average characteristics for all products $k \neq j$ at time t. We use the same set of characteristics as before, namely: car size, car max speed, and car horsepower to weight ratio.
Since we are told to replicate the previous step's results with BLP instruments, note that s_{jtc} is also endogenous by construction. Thus, we use the same set of instruments for *both* endogenous predictors.

Table 9: BLP Instrument Regression

	Dependent variable:	
	δ_{jt}	
	(Baseline)	(Nested)
hp_wt	-42.878^{***}	-13.254^{*}
	(13.467)	(7.691)
size	1.439^{***}	0.079
	(0.380)	(0.338)
sp	0.058^{**}	-0.006
	(0.025)	(0.016)
pr_s	-0.241^{***}	-0.012
	(0.083)	(0.058)
$\log(s_{jtc})$		0.848^{***}
		(0.205)
Constant	-16.352^{***}	-0.565
	(4.602)	(3.938)
Observations	668	668
R^2	-0.539	0.935
Note:	*p<0.1; **p<0.05; ***p<0.01	

Recall the following piecewise definition for nested logit model (own/cross) price elasticities:

$$\epsilon_{jk} = \frac{\partial s_j}{\partial p_k}\frac{p_k}{s_j} = \begin{cases} \alpha p_j(1-\sigma s_{jc} - (1-\sigma)s_j)/(1-\sigma) & \text{if } k = j \\ -\alpha p_k(\sigma s_{jc} + (1-\sigma)s_k)/(1-\sigma) & \text{if } j,k \in c \\ -\alpha p_k s_k & \text{otherwise} \end{cases} \tag{8}$$

Our preferred specification from Table 8 is (Nested 2) and (Nested) from Table 9.
Producing two (2), *readable* 10×10 matrices for this LaTeX write-up was hard. Thus, we report the estimates from Equation (8) in Figures 1 and 2. "Tiles" are color-coded in relation to the coefficient estimate. Coefficient estimates are posted in the center of tiles, and the kind of elasticity (e.g. from Equation (8)) dictates the color of the coefficient estimate text.

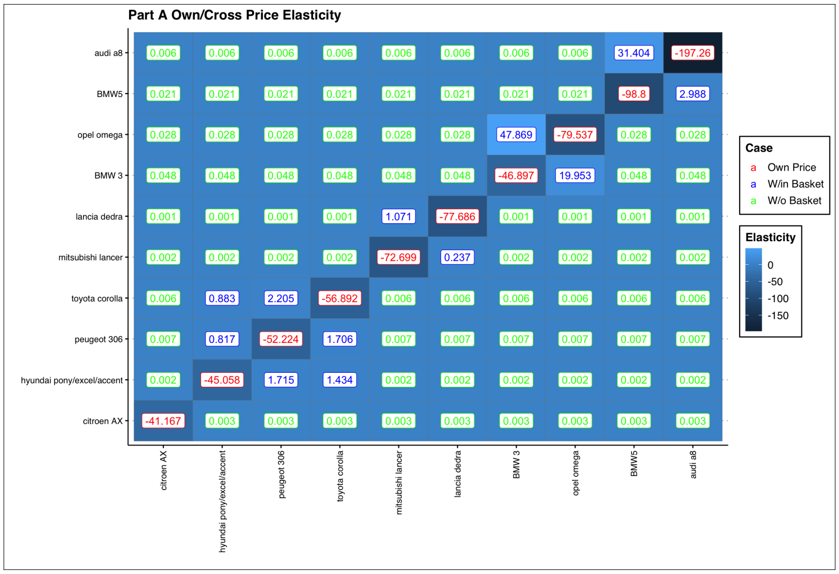

Figure 1: Baseline (Nested) Elasticities

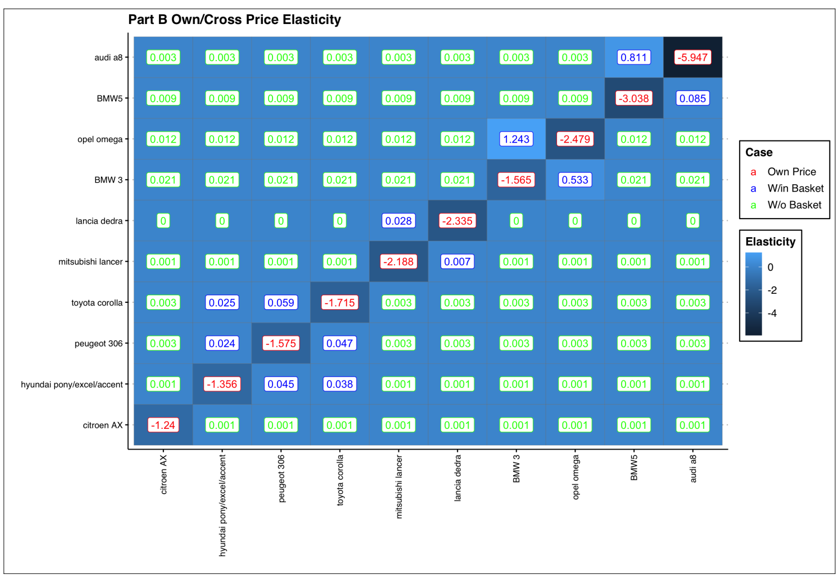

Figure 2: IV (Nested) Elasticities

A big takeaway here is that the baseline (Part A) elasticities are **much** larger than the IV (Part B) elasticities in magnitude terms. Regressing the estimates from Figure 1 on the estimates from Figure 2 suggests that the baseline elasticities are roughly 97% larger (in expectation).

(c) Following Section 3 of Berry, Levinsohn and Pakes (1995), the markup vector for all products $j \in J$ (assuming Bertrand-Nash price competition) is:

$$p - mc = \Delta(p)^{-1} s(p) \tag{9}$$

where $\Delta(p)$ is a $J \times J$ matrix with a typical entry of:

$$\Delta(p)_{jr} = \begin{cases} -\frac{\partial s_r}{\partial p_j} & \text{if products } j \text{ and } r \text{ are produced by the same firm} \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

Since Equation (8) gives us estimates for $\epsilon_{jr} = \frac{\partial s_j}{p_r}\frac{p_r}{s_j}$, then we can populate $\Delta(p)$ from the fact that:

$$-\frac{\partial s_j}{\partial p_r} = -\epsilon_{jr}\frac{s_j}{p_r}$$

We populate elements of $\Delta(p)$ using the parameterizations/results of both the baseline (Part A) and IV (Part B) portions of the exercise. Finally, we **report that Lerner Index**—namely $(p_j - mc_j)/p_j$— **in percentage terms** as a "normalized" measure of markups. Tables 10 and 11 report the top/bottom 5 Lerner index markups. Consistent with Figures 1 and 2, since our own/cross-price elasticities are considerably larger in the baseline relative to the IV model, then the baseline Lernex index values will be correspondingly smaller relative to the IV Lernex index values.

Table 10: Lerner Index (Logit Models) Baseline

	Auto	Base Markup (%)	Price ('000)	HP/Wt	Size	Speed
Top 1	seat marbella	3.70	10.76	0.04	5.21	135
Top 2	fiat panda	3.11	12.69	0.04	5.10	140
Top 3	daihatsu domino/cuore	3.04	12.99	0.05	4.62	135
Top 4	fiat cinquecento	2.98	13.65	0.04	4.79	140
Top 5	opel corsa	2.78	17.70	0.04	6.01	145
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Bot 5	rover 800	0.85	46.30	0.07	8.44	201
Bot 4	saab 9000	0.84	46.95	0.07	8.44	200
Bot 3	alfa 164	0.84	46.95	0.08	8.02	210
Bot 2	mitsubishi sigma	0.76	52.20	0.09	8.43	220
Bot 1	audi a8	0.51	79.60	0.09	9.47	228

Table 11: Lerner Index (Logit Models) IV

	Auto	IV Markup (%)	Price ('000)	HP/Wt	Size	Speed
Top 1	seat marbella	130.24	10.76	0.04	5.21	135
Top 2	fiat panda	103.90	12.69	0.04	5.10	140
Top 3	daihatsu domino/cuore	101.29	12.99	0.05	4.62	135
Top 4	fiat cinquecento	99.33	13.65	0.04	4.79	140
Top 5	opel corsa	92.10	17.70	0.04	6.01	145
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Bot 5	rover 800	28.79	46.30	0.07	8.44	201
Bot 4	citroen XM	28.51	46.53	0.07	8.91	205
Bot 3	164	28.10	46.95	0.08	8.02	210
Bot 2	mitsubishi sigma	25.16	52.20	0.09	8.43	220
Bot 1	audi a8	17.86	79.60	0.09	9.47	228

5.1.3 Logit Model Re-visited

The preceding section used an incorrect definition of market share. I (David) became aware of the mix-up after reading the following Canvas post:

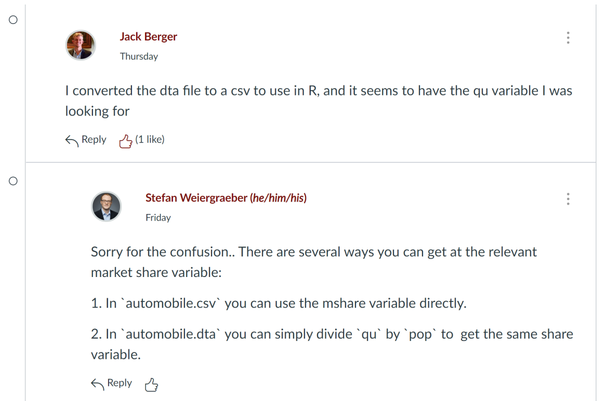

Figure 3: Real market share definition

For this reason, we use `automobile.csv` version of the problem data since this version of the data ensures consistent definition of both market share (e.g. number/unit of sales divided by the revelant market's population) and within category market share.

Moreover, with this sales relative to the population based definition, our calculations for the outside option market share also change. Previously, our outside option market share was defined relative to all other product options to jt, and all observed shares had to sum to unity. In other words, the share of the outside option depended on the share of the current product: the outside option share would change depending on which product we were looking at. This was wrong...

We present corrected versions of the previous section's work below in Table 12.

Corrected Logit Model Results from Table 12

- Column (1) reports the baseline estimation of Equation (7). Note that the price coefficient is positive and significant (as before) due to endogeneity.
- Column (2) adds product j's share of its own nest g (e.g. the nested logit model specification). The price coefficient (α) becomes negative and significant.
- Columns (3) and (4) mirror the specifications of Columns (1) and (2), respectively, and report IV regressions. Column (3) treats price as endogenous whereas Column (4) treats price AND own nest share as endogenous. Columns (3) and (4) use the average horse power to weight ratio, size, and speed for all $jt \neq kt$ as instruments for p_{jt}.

Table 12: Corrected Logit Model Results

	Dependent variable:			
	δ_{jt}			
	OLS		*instrumental variable*	
	(1)	(2)	(3)	(4)
hp_wt	-64.868^{***}	-3.967^{*}	-43.550^{***}	-14.819^{*}
	(7.928)	(2.167)	(13.081)	(7.946)
size	0.355^{***}	0.071^{**}	1.381^{***}	0.062
	(0.116)	(0.030)	(0.369)	(0.350)
speed	-0.013^{*}	-0.002	0.055^{**}	-0.007
	(0.008)	(0.002)	(0.025)	(0.017)
price	0.017^{*}	-0.027^{***}	-0.229^{***}	-0.007
	(0.010)	(0.003)	(0.080)	(0.060)
$\log(s_{j\|g})$		0.978^{***}		0.822^{***}
		(0.010)		(0.212)
Constant	-5.514^{***}	-4.233^{***}	-19.028^{***}	-3.717
	(0.702)	(0.184)	(4.470)	(4.069)
Observations	668	668	668	668
R^2	0.248	0.949	-0.482	0.929
Note:				*p<0.1; **p<0.05; ***p<0.01

For simplicity, we only report the results for the baseline elasticises coming from:

$$\epsilon_{jk} = \frac{\partial s_j}{\partial p_k}\frac{p_k}{s_j} = \begin{cases} \alpha p_j(1-s_j) & \text{if } k = j \\ -\alpha p_k s_k & \text{otherwise} \end{cases} \tag{11}$$

As like before, we report our estimates below in a "tile" plot in Figure 4.

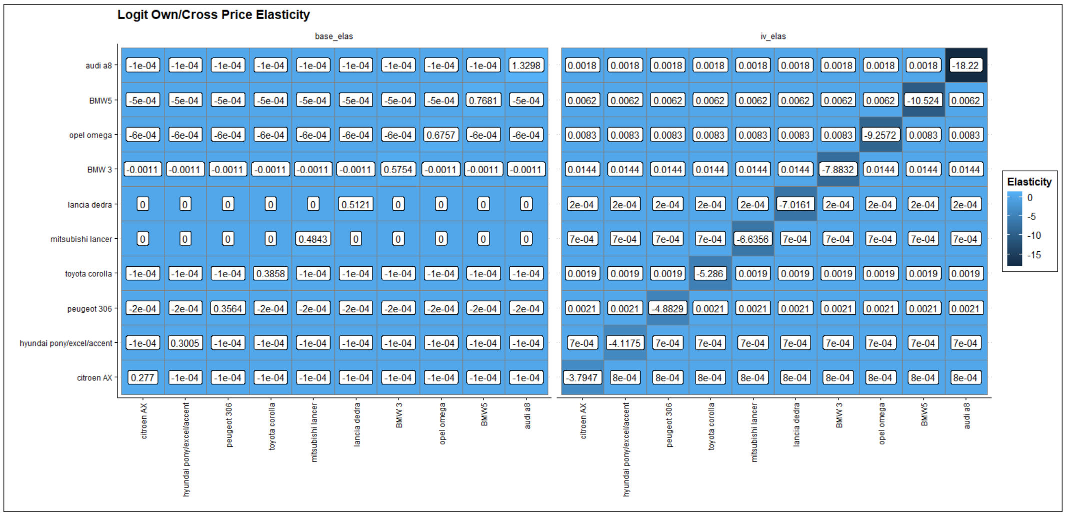

Figure 4: Corrected Logit Elasticities

Any major issues? Yes, there are several (especially with the base model). Since the pricing coefficient was inconsistently estimated...and has a positive sign...this implies both upward sloping demand curve/positive own price elasticities and that all other goods are complements (not substitutes!). Even if we ignore the erroneous signs and just consider magnitudes, we still have issues. With the exception of the Audi A8, all other product own-price elasticities fall in the inelastic region of demand...and standard markup predictions suggest an optimal pricing rule in the elastic portion of the demand curve (unless firms want to receive negative marginal revenue!). In contrast, the IV estimates (e.g. the right hand side panel of Figure 4) seem more "reasonable."

Finally, we recalculate the Lerner Index and report the Top/Bottom 5 markups based on our BLP-instrument logit model (e.g. Column (3) from Table 12).

Table 13: Lerner Index (Logit Model) IV with Corrections

	Model	Lerner Index (%)	Price ('000)	HP/Wt	Size	Speed
Top 1	seat marbella	41.06	10.76	0.04	5.21	135
Top 2	fiat panda	34.47	12.69	0.04	5.10	140
Top 3	daihatsu domino/cuore	33.63	12.99	0.05	4.62	135
Top 4	fiat cinquecento	32.04	13.65	0.04	4.79	140
Top 5	suzuki alto	29.14	14.99	0.05	5.23	150
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Bot 5	saab 9000	9.36	46.95	0.07	8.44	200
Bot 4	alfa 164	9.32	46.95	0.08	8.02	210
Bot 3	mitsubishi sigma	8.37	52.20	0.09	8.43	220
Bot 2	mercedes E klasse	7.97	54.97	0.07	8.25	200
Bot 1	audi a8	5.55	79.60	0.09	9.47	228

5.1.4 Random Coefficients logit model

(a) Our BLP implementation, using the "Nevo" trick, uses:

- Vehicle size and max speed as non-price characteristics with random coefficients (note that price also has a random coefficient);

- Horse power to weight ratio and product dummies as non-price characteristics that are observable to the econometrican; and,
- Average max speed, horse power to weight ratio, and size for car jt (averaging over all cars $k \neq j$ at time t) as instruments.

Table 14: Random coefficients logit model estimates

term	estimate
price	-0.036
hp/wt	-23.953
AlfaRomeo	-6.093
Audi	-2.563
BMW	-2.077
Citroën	-5.304
Daihatsu	-7.059
Fiat	-5.343
Ford	-3.622
Honda	-4.663
Hyundai	-5.778
Lancia	-6.494
Mazda	-4.992
Mercedes	-1.729
Mitsubishi	-4.978
NissanDatsun	-4.861
OpelVauxhall	-3.096
Peugeot	-5.130
Renault	-4.766
RoverTriumph	-6.192
Saab	-5.452
Seat	-5.533
Skoda	-5.265
Suzuki	-6.069
Toyota	-4.637
Volkswagen	-2.992
Volvo	-4.806
Kia	-6.311
$\sigma_p = \beta^u_{price}$	0.00128
$\sigma_{size} = \beta^u_{size}$	0.00010
$\sigma_{speed} = \beta^u_{speed}$	0.00041
GMM Objective	9.6
Computer Time	3.57 minutes

We simulate twenty (20) individuals (e.g. $\nu_{ij} \sim \mathcal{N}(0,1)$ where $i = 20$) for this portion of the exercise (with a fixed seed). Our results are reported in Table 14.

(b) With random coefficients, (own/cross) price elasticity is defined as:

$$\epsilon_{jk} = \frac{\partial s_j}{\partial p_k}\frac{p_k}{s_j} \approx \begin{cases} -\frac{p_j}{s_j}\sum_{i\in ns}(\alpha+\nu_{ip}\sigma_p)s_{ij}(1-s_{ij}) & \text{if } k=j \\ \frac{p_k}{s_j}\sum_{i\in ns}(\alpha+\nu_{ip}\sigma_p)s_{ij}s_{ik} & \text{otherwise} \end{cases} \tag{12}$$

With this definition of elasticity in mind, we report our estimates for each price-decile car model in the tile plot below in Figure 5:

Figure 5: BLP Elasticities

(c) Top 5/Bottom 5 cars:

Table 15: Lerner Index (BLP Model)

	Model	Lerner Index (%)	Price('000)	HP/Wt	Size	Speed
Top 1	seat marbella	26.26	10.76	0.04	5.21	135
Top 2	fiat panda	12.59	12.69	0.04	5.10	140
Top 3	skoda felicia	11.69	15.59	0.04	6.30	145
Top 4	rover mini	11.16	18.30	0.07	4.30	142
Top 5	daihatsu domino/cuore	10.43	12.99	0.05	4.62	135
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Bot 5	BMW5	2.88	46.00	0.06	8.26	198
Bot 4	volvo 850	2.88	46.40	0.07	8.20	215
Bot 3	mercedes C klasse	2.74	43.93	0.07	7.71	193
Bot 2	mitsubishi sigma	2.64	52.20	0.09	8.43	220
Bot 1	mercedes E klasse	2.32	54.97	0.07	8.25	200

5.1.5 Comparison

Table 16 reports various moments of the ratio of BLP to Logit (IV) elasticities. BLP cross-price elasticities are larger than their Logit (IV) counterparts: on average, BLP cross-price elasticities are about 114.5 times larger. This takeway suggests that the BLP model can accommodate greater substitution across different products. Likewise, BLP own-price elasticities tend to run (on average) ≈ 3.3 times larger than their Logit (IV) analogues.

Table 16: BLP/Logit (IV) Elasticities

type	min	q25	median	mean	q75	max
Cross price	0.03	2.63	13.09	114.45	65.26	6454.20
Own price	3.21	3.22	3.25	3.32	3.33	4.28

Figure 6 plots the distribution of pricing coefficients ratios:

$$\frac{\alpha^{BLP} + \nu_{ip}\sigma_p}{\alpha^{Logit}} \quad \forall i = 1, \ldots, ns$$

Figure 7 plots the distribution of markup ratios:

$$\frac{p_j - mc_j^{BLP}}{p_j - mc_j^{Logit}} \quad \forall j = 1, \ldots, J$$

.

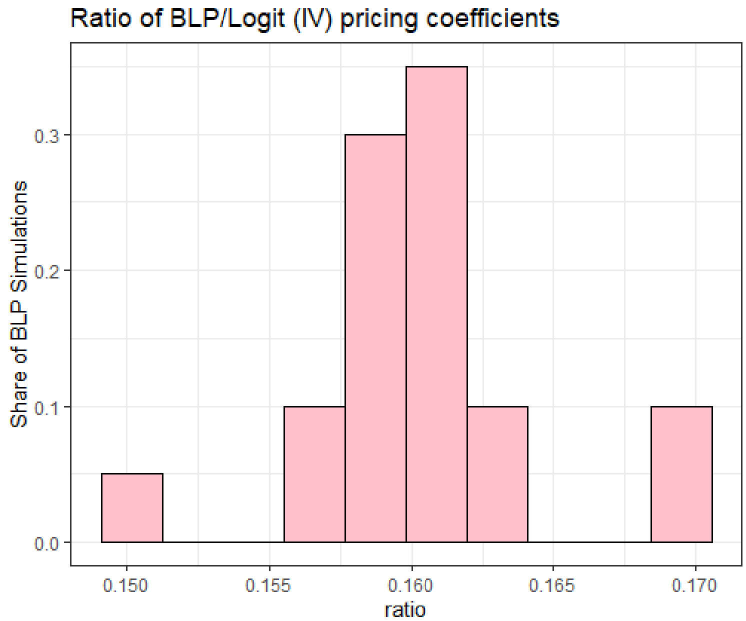

Figure 6: Pricing coefficient ratios

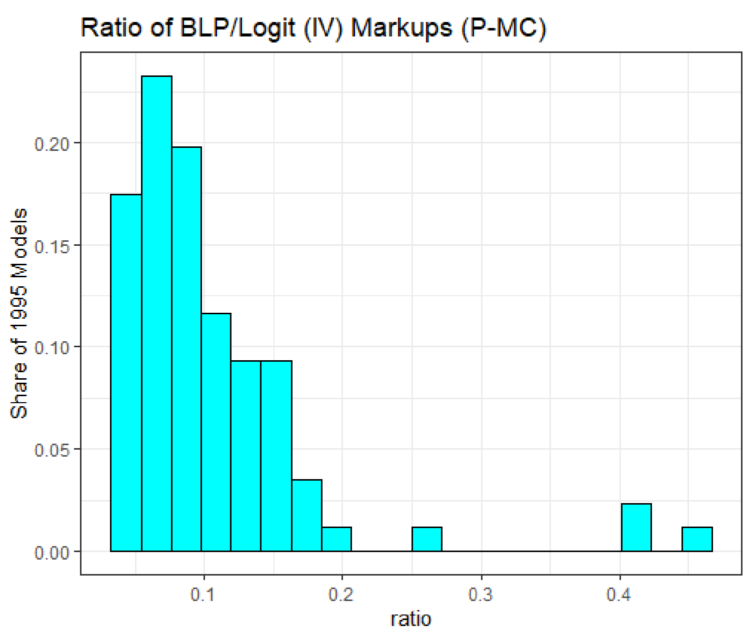

Figure 7: Markup ratios

5.1.6 Income Effect

At this point, we re-present the original work of (Berry, Levinsohn and Pakes, 1995, p.849), focusing on equations (2.7a) and (2.7b). We can re-write utility as:

$$\begin{aligned} u_{ijt} =& \alpha \log(y_{it} - p_{jt}) + x_{jt}\bar{\beta} + \xi_{jt} + \sum_{k=1}^{K} \sigma_k x_{jkt} \nu_{ik} + \epsilon_{ijt} \\ u_{i0t} =& \alpha \log(y_{it}) + \xi_{0t} + \sigma_{0t}\nu_{i0t} + \epsilon_{i0t} \end{aligned}$$

where $x_{jt}\bar{\beta}$ captures the "mean" impact (e.g non-idiosyncratic) of non-price variables on consumer utility whereas terms in red reflect the "random" coefficient aspect of this model. Suppose that income is normally distributed with mean m_t and standard deviation $\hat{\sigma}_{yt}$. Then, following the style of equations (6.14a) and (6.14b) from (Berry, Levinsohn and Pakes, 1995, p.868), we can re-cast our utility function as:

$$\begin{aligned} u_{ijt} =& \alpha \log\Big(\exp\{m_t + \hat{\sigma}_{yt}\nu_{iy}\} - p_{jt}\Big) + x_{jt}\bar{\beta} + \xi_{jt} + \sum_{k=1}^{K} \sigma_k x_{jkt} \nu_{ik} + \epsilon_{ijt} \\ u_{i0t} =& \alpha \log\Big(\exp\{m_t + \hat{\sigma}_{yt}\nu_{iy}\}\Big) + \xi_{0t} + \sigma_{0t}\nu_{i0t} + \epsilon_{i0t} \end{aligned}$$

where vectors $(\nu_{iy}, \nu_{i0}, \ldots, \nu_{ik})$ are drawn from a standard-multivariate distribution prior to estimation. Focusing in on the income term, note that

$$\alpha \log(y_{it} - p_{jt}) = \alpha \log\Big(y_{it}(1 - \frac{p_{jt}}{y_{it}})\Big) = \alpha \log(y_{it}) + \alpha \log\Big(1 - \frac{p_{jt}}{y_{it}}\Big) \approx \alpha \log(y_{it}) - \alpha \frac{p_{jt}}{y_{it}}$$

using a first-order Taylor expansion approximation. We normalize u_{i0t} to be zero (0) by subtracting off $\alpha \log(y_i t) + \sigma_{0t}\nu_{i0t}$ from all terms. Next, if we let $\bar{\alpha} = \alpha / y_{it}$, then

$$\alpha \log(y_{it}) - \frac{\alpha}{y_{it}} p_{jt} = \alpha \log(y_{it}) - \bar{\alpha} p_{jt}$$

Finally, we can redefine our mean utility and deviation from the mean utility terms as

$$\begin{aligned} \delta_{jt} &= -\bar{\alpha} p_{jt} + x_{jt}\bar{\beta} + \xi_{jt} \\ \mu_{ijt} &= \alpha \log(y_{it}) + \sum_{k=1}^{K} \sigma_k x_{jkt} \nu_{ik} \end{aligned}$$

With these two terms in mind, define implied market shares as

$$s_{jt} = \frac{1}{ns} \sum_{i=1}^{ns} \frac{\exp\{\delta_{jt} + \mu_{ijt}\}}{1 + \sum_{l=1}^{J} \exp\{\delta_{lt} + \mu_{ilt}\}}$$

With these shares defined, we can employ the fixed point approach to determine the δ_{jt} given a starting guess of $(\alpha, \sigma_1, \ldots, \sigma_K)$. Once the fixed point "inner-loop" problem is solved—producing δ^*_{jt}— we can identify $\bar{\alpha}$ and $\bar{\beta}$ by regressing δ^*_{jt} on p_{jt} and x_{jt} (instrumenting where needed). The last step is search non-linearly for parameters $(\alpha, \sigma_1, \ldots, \sigma_K)$ to minimize the relevant GMM function.

Our estimation of the substitution matrix is really not that different from any other portion of this assignment. In this case, we have the (j, k)-element of the substitution matrix Δ as

$$\Delta_{j,k} = \begin{cases} -\frac{\partial s_k}{\partial p_j} & \text{if } j, k \text{ produced by same firm} \\ 0 & \text{otherwise} \end{cases}$$

where

$$\frac{\partial s_k}{\partial p_j} = \begin{cases} \int \frac{\partial \delta_{kt}}{\partial p_{kt}} s_{kt}(1-s_{kt})dF(\nu) = \int -\underbrace{\frac{\alpha}{y_{it}}}_{=\bar{\alpha}} s_{kt}(1-s_{kt})dF(\nu) & \text{for } j=k \\ \int -\frac{\partial \delta_{jt}}{\partial p_{jt}} s_{kt}s_{jt}dF(\nu) = \int \underbrace{\frac{\alpha}{y_{it}}}_{=\bar{\alpha}} s_{kt}s_{jt}dF(\nu) & \text{for } j \neq k. \end{cases}$$

5.1.7 Merger Analysis

Equation (9) comes from the first order conditions of the firm's profit maximization problem. The relevant FOC is:

$$s(p) - \Delta(p)(p - mc) = 0$$

where $\Delta(p)$ is defined as before in equation (10). Rather than solving for a mark-up, we can back out a vector of marginal costs for all products:

$$mc = \Delta^{-1}s(p) + p \tag{13}$$

Save this marginal cost vector. Reformulate the first order condition to solve for a new price vector $\hat{p}$ that reflects changes to firm/product shares post merger:

$$\hat{p} = mc - \Delta^{-1}(\hat{p})s(\hat{p}) \tag{14}$$

Equation (14) implies a non-linear system with J equations and J unknowns (from a total of J products). Put differently, we need to employ a non-linear equation approach to solve for the $J \times 1$ price vector satisfies (14).[3]. Other relevant notes:

- mc estimates come from (13), which reflect pre-merger data/estimates.
- Elements of Δ^{-1} depend on post-merger prices. We recast (10) and report a typical element as:

  $$\Delta(p)_{jr} = \Omega_{rj} = -\frac{\partial s_r}{\partial p_j}$$

  Think of Ω_{rj} as a(n) (indicator) function which embeds the structure of market ownership; $\Omega_{rj} = 1$ if products r and j are produced by the same firm. Next, **we assume that pre-merger elasticities stay the same post-merger**. Post merger, we can define $\Delta(\hat{p})_{jr}$ as

  $$\Delta(\hat{p})_{jr} = -\Omega^{post}_{rj}\epsilon_{rj}\frac{s(\hat{p}_j)}{\hat{p}_r}$$

  where Ω^{post}_{rj} reflects the post merger market structure.
- $s(\hat{p})$ will be a bit more involved to define.
  - **Logit**:

$$s(\hat{p}_j) = \frac{\exp\{x_j\beta - \alpha\hat{p}_j + \xi_j\}}{1 + \sum_{l=1}^{J}\exp\{x_m\beta + \alpha\hat{p}_m + \xi_m\}} \tag{15}$$

[3] In practice, we minimize max/sup absolute error implied by Equation (14)

– **BLP**:

$$s(\hat{p}_j) = \frac{1}{ns}\sum_{i=1}^{ns}\frac{\exp\left\{x_j\beta + \left(\sum_{k=1}^{2}\sigma_k\nu_{ik}x_j\right) + (\alpha+\sigma_p\nu_{ip})\hat{p}_j + \xi_j\right\}}{1+\sum_{l=1}^{J}\exp\left\{x_j\beta + \left(\sum_{k=1}^{2}\sigma_k\nu_{ik}x_j\right) + (\alpha+\sigma_p\nu_{ip})\hat{p}_j + \xi_j\right\}} \tag{16}$$

- If we wanted to compute changes in welfare post merger, we could follow Manski, McFadden et al. (1981) and calculate compensating variation according to:

  – **Logit**:

$$CV = \frac{1}{\alpha}\Big(\ln\Big(1+\sum_{j=1}^{J}\exp\{x_j\beta - \alpha\hat{p}_j + \xi_j\}\Big) - \ln\Big(1+\sum_{j=1}^{J}\exp\{x_j\beta - \alpha p_j + \xi_j\}\Big)\Big)$$

  – **BLP**:

$$CV = \frac{1}{ns}\sum_{i=1}^{ns}\frac{1}{\alpha_i}\Bigg[\Bigg(\ln\Bigg(1+\sum_{j=1}^{J}\exp\left\{x_j\beta + \left(\sum_{k=1}^{2}\sigma_k\nu_{ik}x_j\right) + (\alpha+\sigma_p\nu_{ip})\hat{p}_j + \xi_j\right\}\Bigg)\dots$$
$$- \ln\Bigg(1+\sum_{j=1}^{J}\exp\left\{x_j\beta + \left(\sum_{k=1}^{2}\sigma_k\nu_{ik}x_j\right) + (\alpha+\sigma_p\nu_{ip})p_j + \xi_j\right\}\Bigg)\Bigg)\Bigg]$$

Merger Analysis First Attempt

We went ahead and asked the question: "what happens to consumer welfare if each country merges all of its firms"? Moreover, how does the size in welfare change vary by demand model (e.g. BLP vs Logit)? The answer to the first question? Consumer welfare doesn't seem to vary that much pre/post merger exercise. The answer to the second question? Well, BLP merger exercises tend to result in (i) higher CV estimates (relative to Logit)..roughly twice as much welfare gain; and, (ii) lower relative change in prices. How does result (i) square with result (ii)? Our conjecture is that since

$$\frac{\alpha^{BLP} + \nu_{ip}\sigma_p}{\alpha^{Logit}} < 0.17 \quad \forall i = 1,\dots,ns$$

that relatively smaller pricing coefficients $\alpha_i = \alpha^{BLP} + \nu_{ip}\sigma_p$ more than compensate for the relatively smaller change in relative prices.

We present the results of this exercise in two ways.

First, we produce a barplot that depicts compensating variation per country per demand model. Countries are arranged in the order of pre-merger market size (per country). Also, note that since we scaled down our pricing variable by a factor of $1,000$, we should multiply our compensating variation estimates by $1,000$ to arrive a dollar value (or whatever currency denomination corresponds to `pr`is the data). See Figure 8.

Second, we produce boxplots per country per demand model of the ratio of post-merger prices to pre-merger prices. Note that our estimates suggest that *all mergers lower prices* and that *prices fall in greater relative proportion for logit vs BLP models*. See Figure 14.

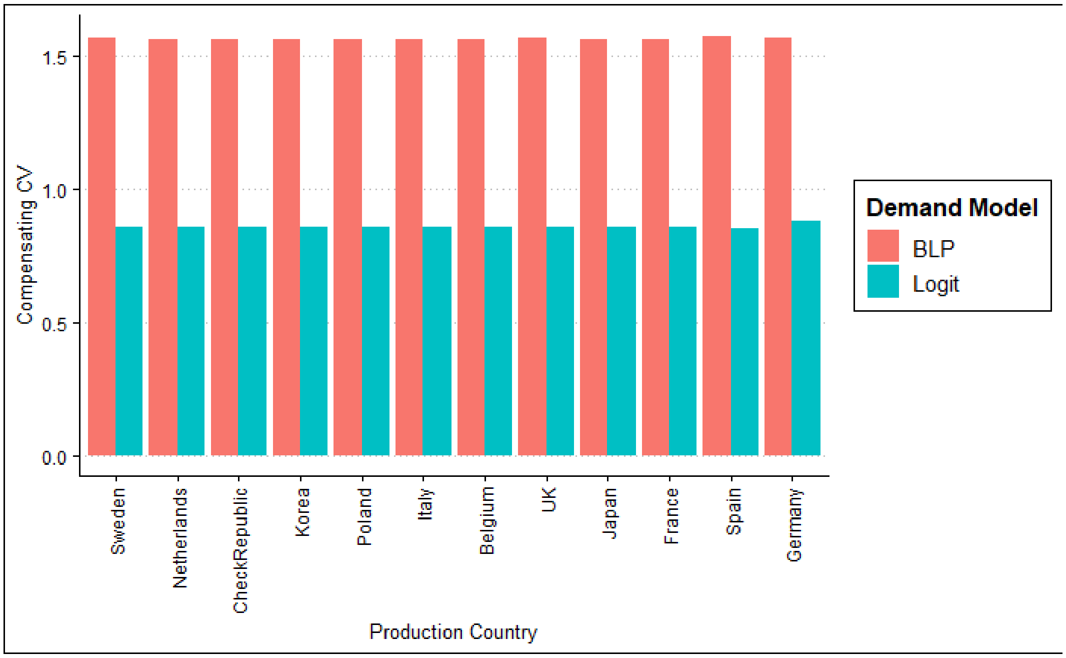

Figure 8: Merger Compensating Variation Estimates

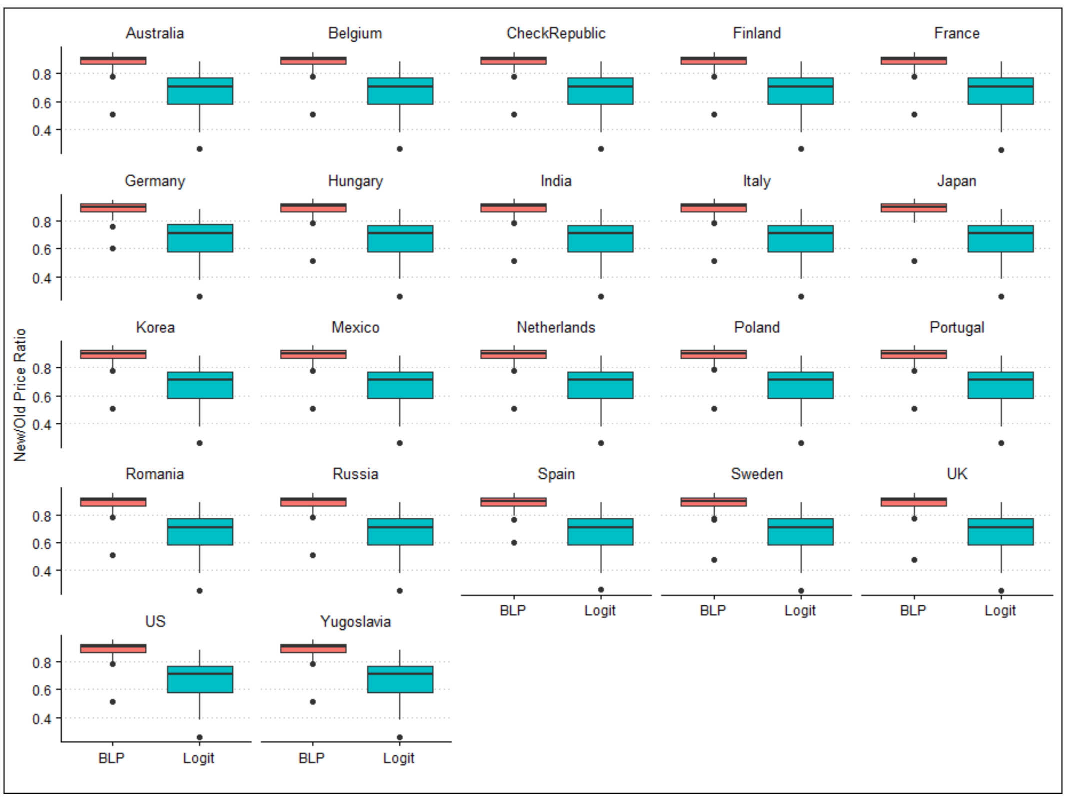

Figure 9: Post/Pre Merger Price Differences

Merger Analysis Second Attempt

Something seemed slightly amiss with our first set of results...namely prices always fell regardless of product or country scenario. It turns out a minus sign was missing at a crucial step of the code, go figure! We re-ran the exercise using both BLP and Logit demand specifications for Germany only. Table 17 reports price changes post merger across both demand-specifications.

Lastly, we compute the change in variable profits as a result of the all German company merger. Focusing exclusively on our BLP demand results, we: (1) assume that pre-merger marginal costs remain constant; (2) assume that total car sales remain the sale pre/post merger; and, (3) use our definition of predicted merger share to estimate post merger sales.

While German produced models tended to make greater post-merger variable profits relative to non-German models (to the tune of 53.8 million on average), some German produced models lost a fair bit of variable profits as well (e.g. German made Mercedes E Klasse and Audi A6 lose 25.2 and 43.4 million, respectively). Figure 10 presents our variable profit results:

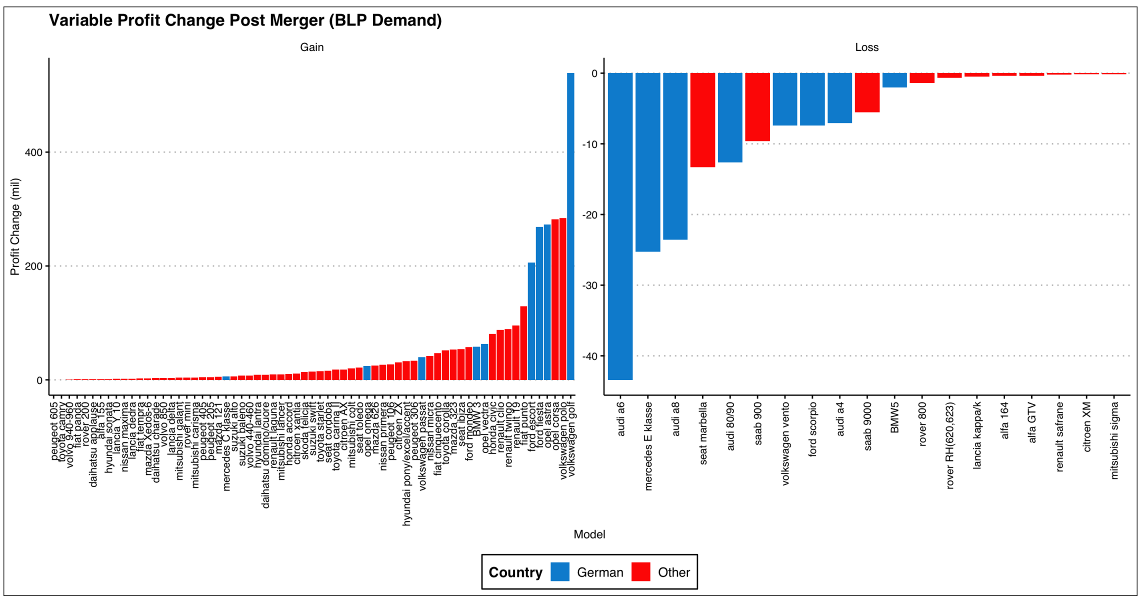

Figure 10: Variable Profit Change

Table 17: German Merger Price Change

		Post Merger Price		Price Change		Price p.p. Change	
Model	p_{old}	Logit	BLP	Logit	BLP	Logit	BLP
audi a8	79.60	79.60	75.44	0.00	-4.16	0.00	-5.23
mercedes E klasse	54.97	55.00	55.41	0.03	0.44	0.05	0.80
mitsubishi sigma	52.20	52.16	52.71	-0.04	0.51	-0.08	0.97
alfa 164	46.95	47.02	46.79	0.07	-0.16	0.14	-0.33
saab 9000	46.95	46.78	44.25	-0.17	-2.70	-0.35	-5.75
citroen XM	46.53	46.37	46.87	-0.16	0.34	-0.35	0.74
volvo 850	46.40	46.37	47.02	-0.03	0.62	-0.07	1.33
rover 800	46.30	46.16	44.18	-0.14	-2.12	-0.30	-4.59
BMW5	46.00	46.03	46.32	0.03	0.32	0.06	0.70
lancia kappa/k	45.15	45.08	45.23	-0.07	0.09	-0.15	0.19
peugeot 605	44.94	45.05	45.27	0.11	0.33	0.25	0.73

Table 17: German Merger Price Change

		Post Merger Price		**Price Change**		**Price p.p. Change**	
Model	p_{old}	Logit	BLP	Logit	BLP	Logit	BLP
volvo 940-960	44.90	44.77	45.38	-0.13	0.48	-0.29	1.07
audi a6	44.90	44.90	44.17	0.00	-0.73	-0.01	-1.63
toyota camry	44.20	44.16	44.52	-0.04	0.32	-0.08	0.72
mercedes C klasse	43.93	43.96	44.25	0.03	0.32	0.06	0.72
alfa GTV	43.80	43.73	43.41	-0.07	-0.39	-0.17	-0.90
renault safrane	40.90	40.86	40.85	-0.04	-0.05	-0.09	-0.11
opel omega	40.48	40.48	40.92	0.00	0.45	0.01	1.10
ford scorpio	40.47	40.49	40.07	0.02	-0.40	0.04	-0.98
audi 80/90	40.45	40.44	39.87	-0.01	-0.58	-0.02	-1.45
nissan maxima	39.99	40.04	40.38	0.05	0.38	0.12	0.96
saab 900	38.95	38.96	35.70	0.01	-3.25	0.03	-8.33
audi a4	37.40	37.39	37.29	-0.01	-0.11	-0.03	-0.30
mazda Xedos-6	36.99	36.90	37.49	-0.09	0.50	-0.23	1.34
mitsubishi galant	36.38	36.34	36.96	-0.04	0.58	-0.11	1.58
BMW 3	34.50	34.53	34.71	0.03	0.21	0.08	0.60
alfa 155	34.10	34.10	34.36	0.00	0.26	0.00	0.78
volkswagen passat	33.50	33.48	33.62	-0.02	0.12	-0.06	0.36
rover RH(620,623)	33.10	33.02	32.53	-0.08	-0.57	-0.23	-1.71
ford mondeo	32.42	32.34	32.70	-0.08	0.28	-0.24	0.87
hyundai sonata	30.99	30.92	31.41	-0.07	0.42	-0.22	1.35
opel vectra	30.95	30.95	31.24	0.00	0.29	0.00	0.95
honda accord	30.88	30.86	31.34	-0.02	0.46	-0.05	1.51
lancia dedra	30.65	30.55	30.95	-0.10	0.30	-0.32	0.97
renault laguna	30.65	30.52	30.55	-0.13	-0.10	-0.43	-0.33
toyota carina II	30.58	30.50	30.90	-0.08	0.32	-0.26	1.04
mazda 626	29.99	29.94	30.59	-0.05	0.60	-0.16	2.01
citroen xantia	29.95	29.95	30.29	0.00	0.34	-0.02	1.12
volvo 440-460	29.90	29.88	30.20	-0.02	0.30	-0.06	1.00
peugeot 405	29.64	29.53	30.12	-0.11	0.48	-0.38	1.63
lancia delta	29.50	29.45	29.89	-0.05	0.39	-0.19	1.34
nissan primera	29.25	29.17	29.78	-0.07	0.54	-0.25	1.84
mitsubishi lancer	28.99	28.98	29.42	-0.01	0.43	-0.05	1.49
mitsubishi carisma	28.99	29.08	29.26	0.09	0.27	0.31	0.93
volkswagen vento	28.08	28.07	27.21	-0.01	-0.87	-0.02	-3.11
hyundai lantra	26.29	26.26	26.74	-0.03	0.45	-0.11	1.71
rover 200	25.25	25.19	24.76	-0.06	-0.49	-0.22	-1.96
fiat tempra	24.57	24.52	24.50	-0.04	-0.07	-0.18	-0.28
daihatsu applause	24.30	24.24	24.71	-0.06	0.41	-0.23	1.69
seat toledo	23.50	23.52	23.48	0.02	-0.01	0.09	-0.06
ford escort	23.18	23.17	23.72	-0.01	0.54	-0.05	2.32
toyota corolla	23.10	23.11	23.55	0.01	0.45	0.03	1.97
mazda 323	22.99	22.97	23.23	-0.02	0.24	-0.09	1.03
opel astra	22.88	22.87	23.23	-0.01	0.35	-0.03	1.54
mazda 121	22.74	22.72	23.22	-0.02	0.48	-0.08	2.10
mitsubishi colt	22.49	22.45	23.00	-0.04	0.51	-0.16	2.27
volkswagen golf	22.43	22.42	23.09	-0.01	0.66	-0.05	2.95
seat cordoba	21.74	21.67	21.44	-0.07	-0.30	-0.32	-1.38

Table 17: German Merger Price Change

		Post Merger Price		**Price Change**		**Price p.p. Change**	
Model	p_{old}	Logit	BLP	Logit	BLP	Logit	BLP
citroen ZX	21.62	21.61	22.19	-0.01	0.57	-0.06	2.62
peugeot 306	21.34	21.27	21.24	-0.07	-0.10	-0.33	-0.46
renault 19	21.25	21.25	21.87	0.00	0.62	-0.01	2.90
honda civic	20.49	20.48	21.00	-0.01	0.51	-0.07	2.51
peugeot 205	19.99	19.96	20.22	-0.03	0.23	-0.16	1.17
suzuki baleno	19.98	19.95	19.97	-0.03	-0.01	-0.17	-0.06
toyota starlet	19.56	19.67	20.07	0.11	0.51	0.55	2.59
daihatsu charade	18.80	18.91	19.12	0.11	0.32	0.60	1.70
rover mini	18.30	18.23	18.11	-0.07	-0.19	-0.36	-1.05
volkswagen polo	18.30	18.27	18.93	-0.03	0.63	-0.14	3.45
hyundai pony/excel/accent	17.99	17.92	18.40	-0.07	0.41	-0.38	2.28
fiat punto	17.94	17.78	18.32	-0.16	0.38	-0.88	2.10
ford fiesta	17.94	17.94	18.53	0.00	0.59	0.01	3.30
opel corsa	17.70	17.57	18.13	-0.13	0.43	-0.75	2.45
renault clio	17.50	17.33	17.89	-0.17	0.39	-0.97	2.23
seat ibiza	17.00	17.04	17.15	0.05	0.16	0.27	0.94
nissan micra	16.75	16.84	16.81	0.10	0.07	0.60	0.41
suzuki swift	16.68	16.72	16.43	0.04	-0.25	0.23	-1.52
citroen AX	16.58	16.62	16.33	0.04	-0.25	0.26	-1.51
lancia Y 10	16.45	16.55	16.14	0.10	-0.31	0.58	-1.88
renault twingo	16.10	16.19	16.34	0.09	0.24	0.57	1.48
peugeot 106	16.09	16.16	15.72	0.07	-0.37	0.44	-2.30
skoda felicia	15.59	15.65	14.88	0.06	-0.71	0.36	-4.53
suzuki alto	14.99	15.09	15.53	0.09	0.53	0.61	3.56
fiat cinquecento	13.65	13.74	13.68	0.09	0.03	0.65	0.19
daihatsu domino/cuore	12.99	13.09	13.10	0.10	0.11	0.73	0.86
fiat panda	12.69	12.78	11.98	0.09	-0.71	0.70	-5.59
seat marbella	10.76	10.80	8.17	0.05	-2.59	0.43	-24.07

References

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile prices in market equilibrium." *Econometrica: Journal of the Econometric Society*, 841–890.

Berry, Steven T. 1994. "Estimating discrete-choice models of product differentiation." *The RAND Journal of Economics*, 242–262.

Manski, Charles F, Daniel McFadden, et al. 1981. *Structural analysis of discrete data with econometric applications.* MIT press Cambridge, MA.

Nevo, Aviv. 2000. "A practitioner's guide to estimation of random-coefficients logit models of demand." *Journal of economics & management strategy*, 9(4): 513–548.

Train, K. 2009. *Discrete Choice Methods with Simulation.* . Second ed., Cambridge University Press.

Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data.* MIT press.