

Modelling with Discretized Continuous Dependent Variable

Online Appendix

Felix Chan*, Laszlo Matyas** and Agoston Reguly**

April 22, 2021

* Curtin University

** Central European University

1 Technical Proofs

1.1 Proof of Proposition 1.

To prove that the distribution function of the directly transferable observations $f(y_{D^{TS}}^{WS})$, in the magnifying method converges to the distribution of the unobserved variable $f(y)$, let us first state the probabilities of the unobserved variable to fall into class C_b^{WS} ,

$$\Pr(y \in C_b^{WS}) = \Pr(y \in s) \int_{c_{b-1}^{WS}}^{c_b^{WS}} f(y) dy.$$

We use the fact that individual i being assigned to a split sample s is independent of i choosing the class with choice value $z_m^{(s)}$. This is satisfied if the questionnaire is randomly assigned to individuals.

For convergence we want to ensure that in each class in the working sample, there are directly transferable observations. Thus, for each split sample the expected number of directly transferable observations is,

$$\begin{aligned} \mathbb{E}(N_b^{WS}) &= \mathbb{E} \left(\sum_{i=1}^N \mathbf{1}_{\{y_i \in C_b^{WS}\}} \right) \\ &= N \Pr(y \in s) \int_{c_{b-1}^{WS}}^{c_b^{WS}} f(y) dy. \end{aligned} \tag{1}$$

We can reformulate Equation (1) by considering the number of observations up to a certain boundary point, rather than the number of observations in a particular class. That is checking for

$$\Pr \left(\mathbb{E} \left[\sum_{i=1}^b N_i^{WS} \right] > 0 \right) \rightarrow 1.$$

This gives the possibility to replace $\int_{c_{b-1}^{WS}}^{c_b^{WS}} f(y)dy$ with $\int_{c_0^{WS}}^{c_b^{WS}} f(y)dy$. Since this is a CDF, and hence a non-decreasing function, which is effectively showing that each class has non-empty observations, we can write the following:

$$\begin{aligned} \mathbb{E} \left(\sum_{i=1}^b N_i^{WS} \right) &= \mathbb{E} \left(\sum_{i=1}^N \mathbf{1}_{\{y_i < c_b^{WS}\}} \right) \\ &= N \Pr(y \in s) \int_{c_0^{WS}}^{c_b^{WS}} f(y)dy. \end{aligned}$$

Next, we need to show that this is an increasing function in C_b^{WS} . Now as $N \rightarrow \infty$, under the assumption that $\Pr(x \in s) = 1/S$ and $S/N \rightarrow c$ with $c \in (0, 1)$ (which is satisfied when $S = cN$)

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E} \left(\sum_{i=1}^b N_i^{WS} \right) &= N \Pr(y_i < C_b^{WS}) \\ &= \frac{1}{c} \int_{C_0^{WS}}^{C_b^{WS}} f(y)dy. \end{aligned}$$

Note that the derivative with respect to C_b^{WS} is $\frac{1}{c} D(C_b^{WS}) > 0$, so the expected number of observations in each class is not 0. This completes our proof.

1.2 Derivation of asymptotic properties of the replacement estimator

Let us formalize the non-directly transferable observations as $y_i^{(s)} \in C_\chi$ and emphasise the non-directly transferable intervals χ , where

$$C_\chi := \zeta^{\mathbb{L}} = \bigcup_{s,m} C_m^{(s)} \bigcap_b C_b^{WS}$$

is the set for non-directly transferable observations from all split samples, with $\chi = 1, \dots, 2(S-1)$. We can then replace $y_i^{(s)} \in C_\chi$ with $\hat{\pi}_\chi$, which denotes the sample conditional averages

$$\hat{\pi}_\chi = \left(\sum_{i=1}^N \mathbf{1}_{\{y_{i,DTO}^{WS} \in C_\chi\}} \right)^{-1} \sum_{i=1}^N \mathbf{1}_{\{y_{i,DTO}^{WS} \in C_\chi\}} y_{i,DTO}^{WS}.$$

Let us call $\hat{\pi}_\chi$ the ‘replacement estimator’ of the conditional expectation of the given class. Under the WLLN, it is straightforward to show that the ‘replacement estimator’ for the sample conditional averages converges to the conditional expectations, thus $\hat{\pi}_\chi \rightarrow \mathbb{E}(y|y \in C_\chi)$ as $N, S \rightarrow \infty$ under the same assumptions as for Proposition 1.

To obtain asymptotic standard errors of this estimator, one can think of $\hat{\pi}_\chi$ as an LS estimator, regressing $\mathbf{1}_{\{y_{i,DTO}^{WS} \in C_\chi\}}$ on $y_{i,DTO}^{WS}$, where $\mathbf{1}_{\{y_{i,DTO}^{WS} \in C_\chi\}}$ is a vector of indicator variables, created by $2(S-1)$ indicator functions. The resulting dummy variables are not independent of each other, while the non-transferable observation classes (C_χ) are overlapping each other. We can now write the following:

$$y_{i,DTO}^{WS} = \boldsymbol{\pi}_\chi \mathbf{1}_{\{y_{i,DTO}^{WS} \in C_\chi\}} + \eta_i,$$

where $\boldsymbol{\pi}_\chi$ stands for the vector of $\pi_\chi, \forall \chi$. The LS estimator of $\boldsymbol{\pi}_\chi$ is

$$\hat{\boldsymbol{\pi}}_\chi = \left(\mathbf{1}'_{\{y_{i,DTO}^{WS} \in C_\chi\}} \mathbf{1}_{\{y_{i,DTO}^{WS} \in C_\chi\}} \right)^{-1} \mathbf{1}'_{\{y_{i,DTO}^{WS} \in C_\chi\}} y_{i,DTO}^{WS},$$

and under the standard OLS assumptions, we can write

$$\sqrt{N_{DTO}^{WS}} (\hat{\boldsymbol{\pi}}_\chi - \boldsymbol{\pi}_\chi) \overset{a}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_\chi),$$

where $\boldsymbol{\pi}_\chi = \mathbb{E}(y|y \in C_\chi), \forall \chi$.

The variance of the LS estimator is given by the usual expression,

$$\boldsymbol{\Omega}_\chi = V(\eta_i) \left(\mathbf{1}'_{\{y_{i,DTO}^{WS} \in C_\chi\}} \mathbf{1}_{\{y_{i,DTO}^{WS} \in C_\chi\}} \right)^{-1}.$$

1.3 Proof of Proposition 2

Let us define the probability of y falling into the working sample’s interval C_b^{WS} ,

$$\Pr(y \in C_b^{WS}) = \sum_{s=1}^S \Pr(y \in s) \sum_{m=1}^M \Pr(y \in C_b^{WS} \mid y \in C_m^{(s)}) \int_{c_{m-1}^{(s)}}^{c_m^{(s)}} f(y) dy. \quad (2)$$

As $S \rightarrow \infty$, $\exists c_b^{WS} = c$ for any $c \in (a_l, a_u)$, by construction. Furthermore, for any c_b^{WS} , $\exists l \in [1, S]$, $m \in [1, M]$ such that $c_b^{WS} = c_m^{(l)}$. Also note that as $S \rightarrow \infty$, we need $N \rightarrow \infty$ as well. Now consider $\Pr(y^\dagger < c_b^{WS}) = \Pr(y^\dagger < c_m^{(l)})$, given $\Pr(y \in S) = 1/S$ and using equation (2) gives

$$\Pr(y^\dagger < c_m^{(l)}) = \frac{1}{S} \sum_{s=1}^S \Pr(y < c_m^{(l)} | y < c_m^{(s)}) \Pr(y < c_m^{(s)}).$$

Note that the summation over the different classes in Equation (2) is being replaced as we are considering the cumulative probability and that no value greater than $c_m^{(l)}$ will be used as a candidate in the working sample for c_b^{WS} . Under the shifting method, $c_m^{(s)} \leq c_m^{(l)}$ for $s < l$ and using the definition of conditional probability gives

$$\begin{aligned} \Pr(y^\dagger < c_m^{(l)}) &= \frac{1}{S} \sum_{s=1}^S \Pr(y < c_m^{(l)}, y < c_m^{(s)}) \\ &= \frac{1}{S} \sum_{s=1}^l \Pr(y < c_m^{(l)}, y < c_m^{(s)}) + \frac{1}{S} \sum_{s=l+1}^S \Pr(y < c_m^{(l)}, y < c_m^{(s)}) \\ &= \frac{1}{S} \sum_{s=1}^l \Pr(y < c_m^{(s)}) + \frac{1}{S} \sum_{s=l+1}^S \Pr(y < c_m^{(l)}). \end{aligned}$$

The last line follows from the fact that $\Pr(y < a, x < b) = \Pr(y < a)$ if $a < b$, and the construction of the shifting method allows us to always disentangle the two cases. Since l is fixed

$$\begin{aligned} \Pr(y^\dagger < c_m^{(l)}) &= \frac{S-l-1}{S} \Pr(y < c_m^{(l)}) + \frac{1}{S} \sum_{s=1}^l \Pr(y < c_m^{(s)}) \\ \lim_{S \rightarrow \infty} \Pr(y^\dagger < c_m^{(l)}) &= \Pr(y < c_m^{(l)}). \end{aligned}$$

This completes the proof.

1.4 Speed of Convergence for the Shifting Method

The probabilities of the artificially created variable y_i^\dagger to be fall into C_b^{WS} bin is given by,

$$\Pr(y_i^\dagger \in C_b^{WS}) = \begin{cases} 0, & \text{if } s = 1 \text{ and } m = 1, \\ \frac{1}{S} \sum_{s=2}^S \frac{1}{s-1} \int_{C_1^{(s)} | C_b^{WS} \in C_1^{(s)}} f(y) dy, & \text{if } s \neq 1 \text{ and } m = 1, \\ \frac{1}{S^2} \sum_{s=1}^S \int_{C_m^{(s)} | C_b^{WS} \in C_m^{(s)}} f(y) dy, & \text{if } 1 < m < M, \\ \frac{1}{S} \sum_{s=1}^S \frac{1}{S-s+1} \int_{C_M^{(s)} | C_b^{WS} \in C_M^{(s)}} f(y) dy, & \text{if } m = M. \end{cases} \quad (3)$$

For each of the conditions in Equation (3), the corresponding expression is $o(1)$. To see this, note that $f(y)$ is a density, so the integral is less than 1. First, consider the case of $s \neq 1$ and $m = 1$,

$$\begin{aligned} \frac{1}{S} \sum_{s=2}^S \frac{1}{s-1} \int_{C_1^{(s)} | C_b^{WS} \in C_1^{(s)}} f(y) dy &\leq \frac{1}{S} \sum_{s=2}^S \frac{1}{s-1} \\ &= \frac{1}{S} \sum_{s=1}^S \frac{1}{s} \\ &= \frac{1}{S} \int_1^S \frac{1}{s} ds \\ &= \frac{\log S}{S}. \end{aligned}$$

As $S \rightarrow \infty$, the ratio in the last line goes to 0. This is expected if the widths of the classes in the working sample go to zero. This is straightforward, while the probability that an observation belongs to a point is 0. The same derivations applies to the case when $m = M$. Now, consider the case of $1 < m < M$,

$$\begin{aligned} \frac{1}{S^2} \sum_{s=1}^S \int_{C_m^{(s)} | C_b^{WS} \in C_m^{(s)}} f(y) dy &\leq \frac{1}{S^2} \sum_{s=1}^S 1 \\ &= \frac{1}{S}, \end{aligned}$$

which also converges to 0 as $S \rightarrow \infty$, but at a faster rate than in the previous cases.

2 Additional Monte Carlo simulation results for Section 3.4

We extend the Monte Carlo simulations in five different ways. The basic setup is the same as in Section 3.4, but we change each time one parameter. All the following tables shows the Monte Carlo average bias (or distortion) of $\hat{\beta}$ from $\beta = 0.5$. In parenthesis we report the Monte Carlo standard deviation of the estimated parameter.

Remark: In case of ‘Set identification’, [†] shows that we can only estimate the lower and upper boundaries for the valid parameter set. We report these bounds subtracted with the true parameter, therefore it should give a (close) interval around zero. For ordered choice model, ^{*} shows we report the distortion from the true β is reported. Ordered probit and logit models’ maximum likelihood parameters do not aim to recover the true β parameter, therefore it is not appropriate to call it bias.

2.1 Moderate sample size

First, we investigate the magnitude of the bias, when the sample size is moderate, namely $N = 1,000$.¹ Table 1 shows the results which is similar to the results with $N = 10,000$ as reported in the paper.

	Normal	Logistic	Log-Normal	Uniform	Exponential	Weibull
Set identification [†]	$[-1.1, 1.15]$ (0.06),(0.07)	$[-1.09, 1.15]$ (0.08),(0.08)	$[-1.09, 1.16]$ (0.07),(0.07)	$[-1.07, 1.17]$ (0.09),(0.09)	$[-1.06, 1.18]$ (0.08),(0.09)	$[-1.09, 1.15]$ (0.05),(0.06)
Ordered probit [*]	0.1978 (0.0810)	0.0690 (0.0797)	0.2138 (0.0827)	0.0181 (0.0763)	0.0965 (0.0795)	0.4484 (0.0908)
Ordered logit [*]	0.6523 (0.1479)	0.3828 (0.1431)	0.6967 (0.1561)	0.2419 (0.1364)	0.4309 (0.1455)	1.2109 (0.1682)
Interval regression	0.0254 (0.0618)	0.0329 (0.0784)	0.0398 (0.0694)	0.0512 (0.0882)	0.0638 (0.0825)	0.0396 (0.0505)
Midpoint regression	0.0209 (0.0643)	0.0293 (0.0786)	0.0310 (0.0733)	0.0453 (0.0895)	0.2029 (0.0426)	0.0275 (0.0526)
Magnifying ($S = 10$)	0.0145 (0.1781)	0.0117 (0.2222)	0.0127 (0.1988)	-0.0184 (0.2538)	0.0757 (0.1023)	0.0330 (0.1358)
Shifting ($S = 10$)	0.0016 (0.0682)	-0.0026 (0.0771)	-0.0031 (0.0696)	-0.0053 (0.0872)	0.0050 (0.0441)	-0.0010 (0.0498)

Table 1: Monte Carlo average bias and standard deviation with moderate sample size, $N = 1,000$

¹We have not decreased our sample size further while for magnifying method in case of $N = 100$ it would mean 10 number effective observations on average.

Shifting method always outperforms the alternatives. Magnifying method gives better results, except in the exponential and weibull cases where it has similar magnitude of bias as the interval regression (exponential case) or the midpoint regression (weibull case). Note that in these cases interval regression and midpoint regression are not superior to the magnifying method. They only outperform magnifying method ‘at random’. As we will show in Table 4 these methods are inconsistent in N , however magnifying method does converge to the true parameter value.

2.2 Symmetric boundaries

Next, we investigate symmetric boundary cases. We set the domain for y_i to $a_l = -3, a_u = 3$ and keep x_i generated in the same way. ϵ_i is generated/truncated such that its lower and upper bound is -2 and 2 . In the normal, logistic and uniform cases it means the lower and upper bounds are -2 and 2 . For the log-normal, exponential and weibull cases we truncate at 4 and subtract 2 from the generated distribution.

	Normal	Logistic	Log-Normal	Uniform	Exponential	Weibull
Set identification [†]	$[-1.11, 1.13]$ (0.02),(0.02)	$[-1.15, 1.10]$ (0.02),(0.02)	$[-1.09, 1.16]$ (0.02),(0.02)	$[-1.07, 1.17]$ (0.03),(0.03)	$[-1.06, 1.19]$ (0.03),(0.03)	$[-1.09, 1.15]$ (0.02),(0.02)
Ordered probit*	0.0890 (0.0252)	0.0029 (0.0243)	0.2085 (0.0262)	0.0158 (0.0234)	0.0986 (0.0241)	0.4461 (0.0295)
Ordered logit*	0.4513 (0.0446)	0.3198 (0.0427)	0.6862 (0.0499)	0.2379 (0.0422)	0.4338 (0.044)	1.2085 (0.0546)
Interval regression	0.0085 (0.022)	-0.0267 (0.0234)	0.0371 (0.0221)	0.0491 (0.0271)	0.0663 (0.0249)	0.0397 (0.0166)
Midpoint regression	0.0070 (0.0211)	0.0240 (0.0242)	0.0362 (0.0216)	0.0490 (0.0273)	0.2077 (0.0128)	0.0314 (0.0157)
Magnifying ($S = 10$)	-0.0323 (0.0606)	-0.0336 (0.0694)	-0.0072 (0.0616)	-0.0332 (0.0781)	0.0213 (0.0333)	0.0066 (0.0417)
Shifting ($S = 10$)	-0.0028 (0.0222)	0.0002 (0.0234)	-0.0004 (0.0209)	-0.0023 (0.0278)	-0.0034 (0.0131)	-0.0008 (0.0153)

Table 2: Monte Carlo average bias and standard deviation with symmetric boundary points: $a_l = -3, a_u = 3$

As we expected the maximum likelihood methods, where they have a closer fit to the assumed distribution the distortion is somewhat smaller in case of ordered probit model². This is the case with the normal and logistic

²Note that ordered probit and logit uses different scaling (depending on the assumed distribution), which results in different parameter estimates. In our case it means ordered

distributions for the disturbance term. However the distortion remains with the same magnitude for all the other mis-specified cases. Magnifying method gives slightly worse results in the normal and logistic cases, but the shifting method performs similarly good.

2.3 Number of choices (M)

Another question is how the number of choices (M) effect the bias. We investigated $M = 3$ case, where questionnaire only defines (known) low-mid-high ranges. In general, the bias increases for the methods. Interesting exceptions are interval regression and midpoint regression, where the results become more volatile: in some cases they give better results, while in other even worse. Shifting method gives fairly accurate estimates.

	Normal	Logistic	Log-Normal	Uniform	Exponential	Weibull
Set identification [†]	$[-1.83, 1.90]$ (0.03),(0.03)	$[-1.85, 1.88]$ (0.03),(0.03)	$[-1.85, 1.89]$ (0.03),(0.03)	$[-1.87, 1.87]$ (0.03),(0.03)	$[-1.89, 1.85]$ (0.03),(0.03)	$[-1.81, 1.93]$ (0.02),(0.02)
Ordered probit*	0.1062 (0.0272)	-0.0220 (0.0266)	0.0197 (0.0278)	-0.1028 (0.0250)	-0.0752 (0.0253)	0.2347 (0.0302)
Ordered logit*	0.5193 (0.0462)	0.3220 (0.0457)	0.3916 (0.0472)	0.1700 (0.0423)	0.2169 (0.0428)	0.7246 (0.0509)
Interval regression	0.0124 (0.0224)	0.0124 (0.0281)	0.0122 (0.0268)	-0.0044 (0.0306)	-0.0243 (0.0280)	-0.0061 (0.0200)
Midpoint regression	0.0336 (0.0233)	0.0168 (0.0274)	0.0229 (0.0267)	-0.0011 (0.0307)	-0.2026 (0.0170)	0.0647 (0.0216)
Magnifying ($S = 10$)	0.0138 (0.0534)	0.0022 (0.0676)	0.0179 (0.0606)	-0.0101 (0.0813)	0.0397 (0.0341)	0.0299 (0.0418)
Shifting ($S = 10$)	-0.0234 (0.0247)	-0.0015 (0.0236)	0.0003 (0.0213)	-0.0094 (0.0292)	-0.0128 (0.0154)	-0.0019 (0.0166)

Table 3: Monte Carlo average bias and standard deviation with small number of choice options, $M = 3$

Also note that with split sampling methods the average bias is within 1 standard deviation, which is not true for the other methods, especially when the underlying distribution is exponential or weibull.

logit has higher average distortions than ordered probit, but this is only matter of scaling. One can map one to the other with the scaling factor, $\hat{\beta}_{probit}^{ML} \approx \hat{\beta}_{logit}^{ML} \times 0.25/0.3989$. This is why we use the term distortion rather than bias for these methods.

2.4 Convergence in N

Table 4 shows the (asymptotic) reduction in the bias with the split sampling methods. We use now only normal distribution's setup for ϵ_i . Here we added 'Magnifying with replacement' method, which uses the magnifying method to get the DTOs, then it uses these DTO values to calculate conditional averages to replace the NDTO values.

As the table suggests, as we increase the number of observations the bias vanishes for the split sampling methods. Also if we increase the number of split samples the bias tend to decrease. It is important to highlight that in the magnifying case the effective number of observation is decreasing in S , therefore if we do not increase N the variance of the estimator is increasing. This shows the trade-off between small sample bias and observing the values more precisely. Based on this table we suggest, in case of magnifying method to use only a moderate number of split samples ($3 - 10$) in case of moderate sample size. For the shifting method there is no such trade-off, however the results are not much better as we increase the number of split samples. It is important to highlight the other methods bias/distortion remains the same as we increase the number of observations, therefore they give inconsistent estimates.

		$N = 1,000$	$N = 10,000$	$N = 100,000$
Set identification [†]		$[-1.1, 1.15]$ $((0.06), (0.07))$	$[-1.1, 1.15]$ $((0.02), (0.02))$	$[-1.1, 1.15]$ $((0.01), (0.01))$
Ordered probit*		0.1978 (0.0810)	0.1971 (0.0256)	0.1968 (0.0080)
Ordered logit*		0.6523 (0.1479)	0.6509 (0.0464)	0.6502 (0.0146)
Interval regression		0.0254 (0.0618)	0.0268 (0.0198)	0.0266 (0.0062)
Midpoint regression		0.0257 (0.0635)	0.0251 (0.0195)	0.0251 (0.0061)
Magnifying only DTO	$S = 3$	-0.0526 (0.0916)	-0.0070 (0.0275)	0.0003 (0.0086)
	$S = 5$	0.0116 (0.1226)	0.0379 (0.0363)	-0.0045 (0.0115)
	$S = 10$	0.0217 (0.1694)	-0.0110 (0.0545)	0.0069 (0.0165)
	$S = 25$	0.0939 (0.2522)	-0.0196 (0.0835)	-0.0074 (0.0276)
	$S = 50$	-0.0761 (0.4768)	-0.0050 (0.1233)	0.0075 (0.0392)
	$S = 100$	0.0382 (0.6889)	0.0175 (0.1781)	-0.0033 (0.0557)
Magnifying with replacement	$S = 3$	-0.0597 (0.0986)	-0.0060 (0.0279)	0.0004 (0.0086)
	$S = 5$	-0.0065 (0.1385)	0.0373 (0.0374)	-0.0040 (0.0115)
	$S = 10$	0.0534 (0.1988)	-0.0103 (0.0575)	0.0066 (0.0165)
	$S = 25$	0.1100 (0.3004)	-0.0165 (0.0947)	-0.0075 (0.0280)
	$S = 50$	-0.1135 (0.5403)	-0.0098 (0.1381)	0.0079 (0.0402)
	$S = 100$	0.0918 (0.8123)	0.0189 (0.2061)	-0.0033 (0.0585)
Shifting	$S = 3$	-0.0038 (0.0629)	-0.0018 (0.0198)	-0.0008 (0.0061)
	$S = 5$	-0.0002 (0.0621)	-0.0024 (0.0194)	-0.0001 (0.0059)
	$S = 10$	0.0024 (0.0603)	-0.0016 (0.0189)	-0.0008 (0.0058)
	$S = 25$	0.0013 (0.0592)	-0.0016 (0.0186)	-0.0007 (0.0058)
	$S = 50$	0.0004 (0.0587)	0.0000 (0.0185)	0.0001 (0.0058)
	$S = 100$	0.0004 (0.0596)	-0.0002 (0.0183)	-0.0003 (0.0056)

Table 4: Bias reduction for split sampling methods: different sample sizes and number of split samples

2.5 Magnifying method with replacement

Finally we report our Monte Carlo experiment with magnifying method using all observations using replacement technique based on DTO. These results are slightly worse than using only DTOs with magnifying method.

	Normal	Logistic	Log-Normal	Uniform	Exponential	Weibull
Base	-0.0043 (0.0551)	0.0719 (0.0719)	0.0645 (0.0645)	0.0813 (0.0813)	0.0345 (0.0345)	0.0433 (0.0433)
$N = 1000$	0.0476 (0.2043)	0.0342 (0.2552)	0.0445 (0.2241)	-0.0021 (0.2895)	0.0868 (0.1223)	0.0648 (0.1565)
Symmetric	-0.0323 (0.0639)	-0.0328 (0.0720)	-0.0081 (0.0645)	-0.0317 (0.0813)	0.0176 (0.0345)	0.0053 (0.0433)
$M = 3$	0.0164 (0.056)	0.0026 (0.0713)	0.0181 (0.0638)	-0.0093 (0.0850)	0.0394 (0.0354)	0.0290 (0.0434)

Table 5: Magnifying all observation with replacement using DTOs

Note that this method can be easily computed with the already collected data to check the ‘robustness’ of the magnifying method.