Extracting and Unifying Semi-Elasticities and Effect Sizes from Studies with Binary Dependent Variables

Geraldine Henningsen* Arne Henningsen†‡

Combining the results from regression analyses in a meta-analysis often proves difficult because differences in the statistical methods and in the units of measurement or the encoding of the variables invalidate a direct comparison of the regression coefficients. In this article, we suggest simple and straightforward methods to extract unified measures for quantifying effects on binary dependent variables that are comparable across different studies even if the studies use different statistical methods, different units of measurement, or different codings of the variables. The suggested effect measures can be applied to continuous, interval-coded, and categorical covariates. We, furthermore, suggest methods to obtain valid approximations of the standard errors of the unified effect measures that can be used, e.g., as weighting factors in a subsequent meta-analysis. We have implemented all suggested methods in the R package urbin that we use to demonstrate the application of our methodology.

Keywords: binary dependent variables, meta-analysis, logit, probit

1. Introduction

Combining the results from different regression analyses in a meta-analysis often proves difficult because differences in the applied estimation methods and differences in the units of measurement or the encoding of the variables of interest invalidate a direct comparison of the regression coefficients. For simple linear regression models, e.g., ordinary least squares regression models, these problems can to some extent be overcome by calculating an 'elasticity' for each continuous covariate of interest at the sample mean, and a relative

*corresponding author. E-mail: `geraldine.henningsen@googlemail.com`.

†University of Copenhagen, Department of Food and Resource Economics, Rolighedsvej 23, 1958 Frederiksberg C, Denmark. Phone: +45-3533-2274, e-mail: `arne@ifro.ku.dk`.

‡Geraldine Henningsen was supported by the 'SAVE-E' project funded by the 'Innovation Fund Denmark' (grant number: 4106-00009B). The authors declare no conflicts of interest. Senior authorship is shared.

effect size of each categorical covariate of interest. These measures indicate by how many percent the dependent variable changes if a continuous covariate increases by one percent or if a categorical covariate changes from a reference category to the category of interest. However, for many non-linear regression models, e.g., generalised linear models, this approach is no longer straightforward and often requires more statistical information than is usually provided in articles, in particular if the user wishes to obtain valid standard errors for those measures.

This article introduces simple and straightforward methods to extract comparable effect measures and their corresponding standard errors from studies with binary and categorical dependent variables.[1] We demonstrate how to derive semi-elasticities for continuous covariates and effect size measures for categorical or ordered covariates from the statistical information usually provided in articles. Furthermore, we demonstrate how to transform and unify differently encoded variables, by showing how to calculate a semi-elasticity of an interval-coded covariate, how to calculate the effect size by turning a continuous covariate into an interval-coded covariate, and how to change the reference category or the grouping of a categorical covariate in order to make effects comparable across different studies. Finally, we introduce a simple and novel way to calculate valid approximations of the standard errors for the derived semi-elasticities and effect size measures in cases where the full variance-covariance matrix of the regression model is unavailable—which we deem to be the standard for most publications.

We demonstrate the application of our methodology by means of a data set on women's labour force participation [1] and the R [2] package urbin [3], in which we have implemented all methods that we suggest in this article.

The article is structured as follows: section two gives a brief introduction to the data set; section three briefly presents the regression methods that we cover in this article; sections four to seven discuss the various approaches for calculating semi-elasticities, effect sizes, and corresponding standard errors; section eight demonstrates how these approaches can be applied to non-binary categorical dependent variables; finally, section nine concludes.

2. Data for empirical example

We use an empirical example based on a data set on women's labour force participation [1] to demonstrate the application of our methodology using R package urbin [3] and to test the validity of the approximated standard errors in cases where the full variance-covariance matrix is unavailable to the user. The data set is available through the R package sampleSelection [4] under the name `Mroz87`.

The data set contains 753 observations on married women and their respective labour force participation in the year 1975, as well as various socio-economic background vari-

[1] The estimation methods covered here are the linear probability model, logistic regression, probit regression, ordinal probit regression, multivariate probit regression, and multinomial logistic regression. We refrain for the time being from estimation methods for discrete choice experiments, e.g., conditional logistic regression, or mixed logistic regression.

ables. In total the data set includes 22 variables. Table 1 provides the summary statistics of the variables in this data set.[2]

In our empirical example, we use the women's labour force participation (`lfp`) as dependent (outcome) variable. Variable `lfp` is a dummy variable, where a '1' represents labour force participation and a '0' represents no labour force participation. We regress this variable on a dummy variable for the presence of children in the household (`kids`), the woman's age in years (`age`), and her years of education (`educ`):

$$\Pr(\texttt{lfp} = 1|\texttt{kids}, \texttt{age}, \texttt{educ}) = f(\texttt{kids}, \texttt{age}, \texttt{educ}) \tag{1}$$

Variable `age` is our primary variable of interest and we use it either as a continuous covariate or as an interval-coded covariate with four intervals: 30–37, 38–44, 45–62, and 63–70 years.

To demonstrate how to apply our methods to regression models where the dependent variable has more than two categories, e.g., ordered probit models and multinomial logistic regressions, and to test approximations of standard errors from estimates derived from these regressions models, we create an additional variable for labour force participation that has three categories: 'no labour force participation' (0 working hours), 'part-time labour force participation' (1–1,300 working hours), and 'full-time labour force participation' (¿1,300 working hours).

We estimate equation (1) with the estimation methods discussed in this article. The regression results provide the variance-covariance matrices of the estimated coefficients so that we can apply the Delta method [5] to calculate approximate standard errors of the calculated effect size measures. We use these standard errors as benchmarks to assess the quality of various approximations for cases where all off-diagonal elements of the variance-covariance matrix are unknown, which is the case for most studies published in the literature, as usually only the standard errors (or t-values) of the estimates are reported.

3. Estimation methods

Most estimation methods that can handle binary or categorical dependent variables can be categorised into two groups: methods where the link function follows a normal distribution, so-called probit regressions, and methods where the link function follows a logistic distribution, logistic regressions.[3] Another approach that has regained popularity in recent years because it is based on fewer assumptions than other approaches is the linear probability model, which uses a simple linear link function, and, thus, can be

[2] A more detailed description of the variables in this data set is available, e.g., in the documentation of the sampleSelection package. The R code that loads the data set and prepares it for the examples in Sections 4 to 8 is available in Appendix Section C.1.

[3] There exists a multitude of estimation techniques for models with categorical dependent variables, quasi-categorical dependent variables, like count data, or outcome variables that can be transferred into a binary or categorical variables, like truncated variables. We consider these regression models to be outside the scope of this article.

Table 1: Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Max
lfp	753	0.57	0.50	0	1
hours	753	740.58	871.31	0	4, 950
kids5	753	0.24	0.52	0	3
kids618	753	1.35	1.32	0	8
age	753	42.54	8.07	30	60
educ	753	12.29	2.28	5	17
wage	753	2.37	3.24	0.00	25.00
repwage	753	1.85	2.42	0.00	9.98
hushrs	753	2, 267.27	595.57	175	5, 010
husage	753	45.12	8.06	30	60
huseduc	753	12.49	3.02	3	17
huswage	753	7.48	4.23	0.41	40.51
faminc	753	23, 080.59	12, 190.20	1, 500	96, 000
mtr	753	0.68	0.08	0.44	0.94
motheduc	753	9.25	3.37	0	17
fatheduc	753	8.81	3.57	0	17
unem	753	8.62	3.11	3	14
city	753	0.64	0.48	0	1
exper	753	10.63	8.07	0	45
nwifeinc	753	20.13	11.63	−0.03	96.00
kids	753	1.59	1.46	0	8
age30.37	753	0.33	0.47	0	1
age38.44	753	0.23	0.42	0	1
age45.52	753	0.30	0.46	0	1
age53.60	753	0.13	0.34	0	1

estimated by ordinary least squares (OLS) and other estimators for linear regression models.

The linear probability model assumes that a Bernoulli trial can be explained by a linear combination of covariates:

$$\Pr(Y = 1|X = x) = \beta_0 + \sum_{j=1}^{K} \beta_j x_j, \tag{2}$$

where $Y \in \{0, 1\}$ is a binary dependent variable, $x = (x_1, \ldots, x_K)^\top$ is a vector of K covariates, and $\beta = (\beta_0, \ldots, \beta_K)^\top$ is a vector of $K + 1$ unknown coefficients.

A probit regression model [6, 7] models the same relationship assuming a probit link function which follows a standard normal distribution:

$$\Pr(Y = 1|X = x) = \Phi\left(\beta_0 + \sum_{j=1}^{K} \beta_j x_j\right), \tag{3}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

The logistic regression [8] uses a logit link function:

$$\Pr(Y = 1|X = x) = \frac{\exp\left(\beta_0 + \sum_{k=1}^{K} \beta_k x_k\right)}{1 + \exp\left(\beta_0 + \sum_{k=1}^{K} \beta_k x_k\right)}, \tag{4}$$

A bivariate or multivariate probit model generalises the probit regression model (3) to simultaneously estimate two or more probit equations for different binary dependent variables $Y_1, \ldots, Y_N$, where a potential correlation between the error terms of the different probit equations is explicitly modelled. As meta-analyses usually focus on one specific dependent variable, the coefficients of the regression equations for the other dependent variables and the correlation structure of the error terms in the multivariate probit regression model can be ignored. Hence, the estimation result for the one probit equation of interest can be treated like an estimation result from a univariate probit regression, so that equation (3) can be used to calculate the *unconditional* probabilities $P(Y_n = 1|x_1, \ldots, x_K)$ and the marginal effects on the *unconditional* probabilities in bivariate and multivariate probit regression models [9].[4]

If a study reports the marginal effects based on the regression results of a probit model, a logistic regression, or a multivariate probit model, one can assume a first-order Taylor series approximation of these models around the point, at which the marginal effects were calculated. Under this linear approximation, the marginal effects can be treated as if they were coefficients of a linear probability model (2).

Many extensions of probit and logistic regression models have been developed to accommodate for more complicated data structures. For instance, regression methods for

[4]We do not take into account the conditional probabilities $P(Y_n = 1|x_1, \ldots, x_K, Y_1, \ldots, Y_{n-1}, Y_{n+1}, \ldots, Y_P)$ and the marginal effects on the conditional probabilities, because we focus on one dependent variable and disregard interrelations between different dependent variables.

dependent variables with more than two categories estimate the probability that the dependent variable is equal to a certain category. Studies with this set-up can still be compared to studies with a binary dependent variable if the categories can be grouped into two groups that correspond to the two outcomes of the binary dependent variable in the other studies.

A modification of the probit regression that handles ordered categorical variables, i.e., categorical variables where the ordering of the categories is meaningful (think of first place, second and third place), is the ordered probit regression [10], where the dependent variable can have P distinct and strictly ordered values ($Y \in \{1, \ldots, P\}$), can be specified as:

$$\Pr(Y = p|X = x) = \Phi\left(\mu_p - \sum_{j=1}^{K} \beta_j x_j\right) - \Phi\left(\mu_{p-1} - \sum_{j=1}^{K} \beta_j x_j\right) \quad \forall\, p = 1, \ldots, P, \tag{5}$$

where $\mu_0 < \mu_1 < \ldots < \mu_P$ are the break points, of which $\mu_0 = -\infty$, $\mu_P = \infty$, and $\mu_1, \ldots, \mu_{P-1}$ are unknown and, thus, need to be estimated. To make estimates from an ordered probit model comparable to estimates from models with a binary dependent variable, we create a new binary dependent variable Y^* by dividing the P distinct values of the dependent variable Y into two categories:

$$Y^* = \begin{cases} 0 & \text{if } Y \in \{1, \ldots, p^* - 1\} \\ 1 & \text{if } Y \in \{p^*, \ldots, P\} \end{cases} \tag{6}$$

with $p^* \in \{2, \ldots, P\}$. This reduces the ordered probit model to a binary probit model:[5]

$$\Pr(Y^* = 1|X = x) = \Pr(Y \in \{p^*, \ldots, P\}|X = x) \tag{7}$$

$$= \Phi\left(-\mu_{p^*-1} + \sum_{j=1}^{K} \beta_j x_j\right), \tag{8}$$

where the intercept of the binary probit model (3) is equal to the negative value of the break point that separates $Y^* = 0$ from $Y^* = 1$, i.e., $\beta_0 = -\mu_{p^*-1}$.[6]

The multinomial logistic regression [11] handles estimation models with multinomial dependent variables, i.e., categorical variables where the ordering of the categories has

[5] A proof is given in Appendix Section B.

[6] If the ordered probit model (5) is estimated with intercept, say, β_0^* and (for identification) by normalising the first (internal) break point to zero, i.e., $\mu_1 = 0$, the ordered probit model can be simplified to a binary probit model with intercept $\beta_0 = \beta_0^* - \mu_{p^*-1}$.

no meaning (think of red cars, green, cars, and blue cars):

$$\Pr(Y = p|X = x) \equiv \pi_p \tag{9}$$

$$= \frac{\exp\left(\beta_{0,p} + \sum_{k=1}^{K} \beta_{k,p} x_k\right)}{1 + \sum_{o \in \{1,\ldots,P\} \setminus p^*} \exp\left(\beta_{0,o} + \sum_{k=1}^{K} \beta_{k,o} x_k\right)} \tag{10}$$

$$= \frac{\exp\left(\beta_{0,p} + \sum_{k=1}^{K} \beta_{k,p} x_k\right)}{\sum_{o=1}^{P} \exp\left(\beta_{0,o} + \sum_{k=1}^{K} \beta_{k,o} x_k\right)} \tag{11}$$

$$\forall\, p = 1, \ldots, P,$$

where $Y \in \{1, \ldots, P\}$ is a categorical dependent variable with reference category p^*, $\beta_{\cdot,p} = (\beta_{0,p}, \ldots, \beta_{K,p})^\top; p \in \{1, \ldots, P\} \setminus p^*$ are $P-1$ vectors of $K+1$ unknown coefficients each, and $\beta_{j,p^*} \equiv 0 \;\forall\, j = 0, \ldots, K$ are the $K+1$ coefficients of the reference category p^*, which are all normalized to zero.

4. Semi-elasticities of continuous covariates

4.1. Semi-elasticities

Differences in the units of measurement of the variables of interest often render it impossible to directly compare coefficient estimates from different studies. In many cases, this problem can be circumvented by calculating the elasticity of the effect, $\epsilon_k = \partial \ln(Y)/\partial \ln(x_k) = (\partial Y/\partial x_k) \cdot (x_k/Y)$, e.g., calculated at the sample mean $x_k = \bar{x}_k$ and $Y = \bar{Y}$ (e.g., see [12]). The elasticity describes the percentage change in the dependent variable Y given a one percent increase in the continuous covariate x_k. It is as such unit-free, which allows the user to compare the effect of a particular covariate across different studies.

In the case of studies with binary dependent variables, the regression model describes the probability that the dependent variable has a value of one. As the probability is already coded between zero and one and unit-free, we suggest to calculate a semi-elasticity of each continuous covariate of interest:

$$\epsilon_k \equiv \frac{\partial \Pr(Y = 1|X = x)}{\partial x_k} \cdot x_k, \tag{12}$$

which can be interpreted as the percentage point change in the probability of Y being equal to one given a one percent increase in x_k.

Table 2 presents the equations for calculating the semi-elasticities of continuous covariates for all six estimation methods covered in this article. If the covariate of interest enters the estimation equation both as linear term and as quadratic term, the equations for calculating the semi-elasticities must be extended accordingly. These extensions for quadratic terms are included in the equations in Table 2.

Table 2: Semi-elasticities of continuous covariates for 6 different regression models

Regression model	Semi-elasticity
Linear probability model	$\epsilon_k = (\beta_k + 2\,\beta_{k+1} x_k)\, x_k$
Binary, bivariate / multivariate, and ordered probit regression	$\epsilon_k = \phi\left(\beta_0 + \sum_{j=1}^{K} \beta_j x_j\right) (\beta_k + 2\,\beta_{k+1} x_k)\, x_k$
Logistic regression	$\epsilon_k = \frac{\exp\left(\beta_0 + \sum_{j=1}^{K} \beta_j x_j\right)}{\left(1 + \exp\left(\beta_0 + \sum_{j=1}^{K} \beta_j x_j\right)\right)^2} (\beta_k + 2\,\beta_{k+1} x_k)\, x_k$
Multinomial logistic regression	$\epsilon_{k,\mathscr{P}} = \sum_{p \in \mathscr{P}} \epsilon^*_{k,p}$ with $\epsilon^*_{k,p} = \pi_p \left(\beta_{k,p} + 2\,\beta_{k+1,p} x_k - \sum_{o=1}^{P} (\beta_{k,o} + 2\,\beta_{k+1,o} x_k)\, \pi_o\right) x_k$ $\forall\, p = 1, \ldots, P$

Note: $\phi(\cdot)$ denotes the probability density function of the standard normal distribution. If the regression analysis includes a quadratic term of the covariate of interest, we assume that the model is specified with $x_{k+1} \equiv x_k^2$ so that β_{k+1} and $\beta_{k+1,j}; j = 1, \ldots, P$ indicate the coefficient(s) of the quadratic term of the covariate of interest. If the regression analysis does not include a quadratic term of the covariate of interest, the equations in this table can be simplified by setting all β_{k+1} and all $\beta_{k+1,j}; j = 1, \ldots, P$ to zero. For multinomial logistic regression models, $\mathscr{P}$ indicates the set of categories of the dependent variable that correspond to a binary outcome of one, while all categories that are not in $\mathscr{P}$ correspond to a binary outcome of zero.

As described in Section 3, certain estimates from bivariate, multivariate, and ordered probit regressions can be extracted so that parts of these models correspond to binary probit models and the equation for calculating semi-elasticities of binary probit models can be also applied to bivariate, multivariate, and ordered probit models.

We have implemented the calculation of semi-elasticities for all six estimation methods covered in this article in the R package `urbin`. In order to demonstrate how to use this package, we calculate the semi-elasticity of the variable `age` with regard to a married woman's probability to participate in the labour force based on a probit regression (3) of equation (1).[7]

Table 3: Probit regression results with age as linear and quadratic covariate

	Dependent variable:	
	lfp	
	(1)	(2)
Constant	0.09	-3.89^{***}
	(0.44)	(1.39)
kids	-0.13^{***}	-0.15^{***}
	(0.04)	(0.04)
age	-0.02^{***}	0.17^{***}
	(0.01)	(0.06)
I(age^2)		-0.002^{***}
		(0.001)
educ	0.10^{***}	0.10^{***}
	(0.02)	(0.02)
Observations	753	753
Log Likelihood	-493.99	-489.38
Akaike Inf. Crit.	995.98	988.76
Note:	*p<0.1; **p<0.05; ***p<0.01	

The results of the probit regression are presented in Table 3, which in combination with Table 1 represents the standard information that a user usually can obtain from most publications.

In the following command, function `urbinEla` calculates the semi-elasticity of variable `age`:

[7]The R code for estimating these models is available in Appendix Section C.2.

```r
urbinEla( coef(estProbit), xMean, xPos = 3, model = "probit" )

##      semEla       stdEr
## -0.3608258          NA
```

This is done based on the vector of coefficients (including intercept) of the probit regression, `coef(estProbit)`, and the vector of sample means for all covariates (including a one for the intercept), `xMean`. Argument `xPos` indicates the position of the covariate(s) of interest, in our example `age`, in the vectors `coef(estProbit)` and `xMean`. Argument `model` is set to `"probit"`, because the coefficients are obtained from a probit regression and, thus, the semi-elasticity of variable `age` has to be calculated based on equation (3). The calculated semi-elasticity indicates that the probability that a woman is in the labour force decreases, *ceteris paribus*, by 0.36 percentage points if her age increases by one percent.

If variable `age` also enters the regression equation in quadratic form, we can simply use argument `xPos` to point out the positions of both `age` and `age`^{2} in vectors `coef(estProbitQ)` and `xMeanQ`:

```r
urbinEla( coef(estProbitQ), xMeanQ, xPos = c( 3, 4 ),
  model = "probit" )

##      semEla       stdEr
## -0.3330041          NA
```

If argument `xPos` has two elements, function `urbinEla` automatically uses the extended formula to accommodate the quadratic term in the calculation of the semi-elasticity. The semi-elasticity based on the probit model with both a linear and a quadratic term of `age` indicates that the probability that a woman is in the labour force decreases, *ceteris paribus*, by 0.33 percentage points if her age increases by one percent.

4.2. Approximation of standard errors

An approximate standard error of the semi-elasticity defined in equation (12) can be obtained by using the Delta method [5]:

$$\mathrm{se}\left(\epsilon_k\right) = \sqrt{\frac{\partial \epsilon_k}{\partial \boldsymbol{\beta}} \mathrm{Var}\left(\boldsymbol{\beta}\right) \frac{\partial \epsilon_k}{\partial \boldsymbol{\beta}^\top}}, \tag{13}$$

where $\mathrm{se}\left(\epsilon_k\right)$ indicates the (approximate) standard error of the semi-elasticity ϵ_k, $\partial \epsilon_k / \partial \boldsymbol{\beta}$ indicates the gradient vector of the semi-elasticity ϵ_k with respect to the coefficients $\beta_0, \ldots, \beta_K$, and $\mathrm{Var}\left(\boldsymbol{\beta}\right)$ indicates the variance-covariance matrix of the estimated coefficients. The gradient vectors for the semi-elasticities, $\partial \epsilon_k / \partial \boldsymbol{\beta}$, of the various regression models are presented in Appendix Section A.1.

The following commands calculate the same semi-elasticities as above, but this time include their respective standard errors based on the full variance-covariances matrices of the estimates:

```
urbinEla( coef(estProbit), xMean, xPos = 3, model = "probit",
  allCoefVcov = vcov(estProbit)  )

##     semEla      stdEr
## -0.3608258  0.1145625

urbinEla( coef(estProbitQ), xMeanQ, xPos = c( 3, 4 ),
  model = "probit", allCoefVcov = vcov(estProbitQ) )

##     semEla      stdEr
## -0.3330041  0.1104025
```

As scientific publications usually do not report covariances between estimated coefficients, but only (at best) standard errors (or t-values, which can be used to calculate the standard errors), the covariances between the estimates of the coefficient are usually unknown. A simple solution would be to replace the unknown covariances by zeros. However, in many empirical examples and a few Monte-Carlo trials, we noticed that ignoring the covariances between the coefficients often gives very imprecise, mostly upward-biased, standard errors of the semi-elasticities, particularly if the models include a quadratic term of the covariate of interest:

```
urbinEla( coef(estProbit), xMean, xPos = 3, model = "probit",
  allCoefVcov = sqrt(diag(vcov(estProbit))),
  seSimplify = FALSE  )

## Warning in urbinEla(coef(estProbit), xMean, xPos = 3, model = "probit",
:  the returned standard error is likely very imprecise; you can provide
the full covariance matrix via argument 'allCoefVcov' or do NOT set
argument 'seSimplify' to FALSE to obtain a more precise standard error

##     semEla      stdEr
## -0.3608258  0.1378307

urbinEla( coef(estProbitQ), xMeanQ, xPos = c( 3, 4 ),
  model = "probit", allCoefVcov = sqrt(diag(vcov(estProbitQ))),
  seSimplify = FALSE )

## Warning in urbinEla(coef(estProbitQ), xMeanQ, xPos = c(3, 4), model =
"probit", :  the returned standard error is likely very imprecise; you
can provide the full covariance matrix via argument 'allCoefVcov' or do
NOT set argument 'seSimplify' to FALSE to obtain a more precise standard
error
```

```
## Warning:  In urbinEla(allCoef = coef(estProbitQ), allXVal = xMeanQ,
xPos = c(3, 4), model = "probit", allCoefVcov = sqrt(diag(vcov(estProbitQ))),
seSimplify = FALSE) :
##  the returned standard error is likely largely upward biased and,
thus, in most cases meaningless; you can provide the full covariance
matrix via argument 'allCoefVcov' to avoid this bias or use argument
'xMeanSd' to substantially reduce this bias

##     semEla      stdEr
## -0.3330041  1.7946071
```

In these empirical examples and Monte Carlo trials, we found that when covariances are assumed to be zero, simplifying the calculations of the gradients $\partial \epsilon_k / \partial \boldsymbol{\beta}$ by assuming that the 'weighting factors' in the equations for calculating the semi-elasticities[8] do not depend on the coefficients (although they actually do), gives much better approximations of the standard error than using the correctly calculated gradients. These simplified gradient vectors are presented in Appendix Section A.2. As several elements of these simplified gradient vectors are zero, the calculation of the semi-elasticities with the Delta method ignores many of the unknown covariances so that a lack of covariances causes a smaller problem when using the simplified gradients than when using the full gradients.

The huge overestimation of the standard errors of the semi-elasticities in the presence of a quadratic term of the covariate of interest originates from the multicollinearity between the quadratic term, the linear term, and the intercept. In an OLS regression, e.g., a linear probability model, with an intercept and a linear and quadratic term of a covariate, the variance-covariance matrix of the estimates would be equal to $\sigma^2 (X^\top X)^{-1}$, where σ^2 is the variance of the error term and X is an $N \times 3$ matrix with N the number of observations and its three columns being the intercept and the linear and quadratic term of the covariate, respectively. If we have the values of the covariate, we can calculate the elements $w_{ij}; i, j \in \{1, 2, 3\}$ of the 3×3 matrix $W \equiv (X^\top X)^{-1}$. If we additionally have the standard errors of the coefficients of the linear and quadratic terms of the covariate, i.e., $\text{se}(\beta_1)$ and $\text{se}(\beta_2)$, respectively, we can calculate the variance of the error term as $\sigma^2 = \text{se}(\beta_1)^2 / w_{22}$ or as $\sigma^2 = \text{se}(\beta_2)^2 / w_{33}$ and then the covariance between the two coefficients of the covariate as $\text{Cov}(\beta_1, \beta_2) = \sigma^2 w_{23} = \sigma^2 w_{32}$. As one usually does not have the original data that were used in published studies, but rather the mean value and corresponding standard deviation of the covariate of interest, one can simulate the values of the covariate, e.g., with a pseudo-random number generator sampling from a normal distribution using the actual mean and standard deviation of the covariate. In cases where the covariate is simulated, the actual model includes further covariates, or the actual model is not an OLS model (but, e.g., a probit or logit regression), the two above-described equations for calculating the variance of the error term give two different values. In these cases, one can calculate the approximate error variance as a geometric

[8] I.e., $\phi(\cdot)$ for different types of probit models, $\exp(\cdot)/(1+\exp(\cdot))^2$ for logistic regression models, and π_p and π_o for multinomial logistic regression models, see Table 2.

mean: $\sigma^2 \approx \sqrt{(\text{se}\,(\beta_1)^2/w_{22})\,(\text{se}\,(\beta_2)^2/w_{33})}$, which is a more conservative measure than the arithmetic mean.

If a user of the `urbinEla` function provides standard errors of the coefficients (rather than the full covariance matrix), it uses the simplified gradients to calculate the standard errors unless the user sets argument `seSimplify` to `FALSE`. Moreover, if the model includes a quadratic term of the covariate of interest and the user provides the mean value and the standard deviation of the covariate of interest through argument `xMeanSd`, `urbinEla` uses a pseudo random number generator to draw 1,000 values from a normal distribution with the provided mean value and standard deviation of the covariate and then imputes the covariance between the coefficients of the linear and quadratic term of the covariate as described in the previous paragraph.

The following command uses the simplified gradient and—for the probit regression with the quadratic term—additionally an imputed value of the covariance between the coefficients of the linear and quadratic term of the `age` variable to calculate approximate standard errors of the semi-elasticities:

```
urbinEla( coef(estProbit), xMean, xPos = 3, model = "probit",
  allCoefVcov = sqrt(diag(vcov(estProbit))) )

##     semEla      stdEr
## -0.3608258  0.1145860

urbinEla( coef(estProbitQ), xMeanQ, xPos = c( 3, 4 ),
  model = "probit", allCoefVcov = sqrt(diag(vcov(estProbitQ))),
  xMeanSd = c( mean(Mroz87$age), sd(Mroz87$age) ) )

##     semEla      stdEr
## -0.3330041  0.1333182
```

These standard errors are *much* closer to the standard errors based on the full variance-covariance matrices than the naïve calculations with the full gradients and replacing the missing covariances by zeros.

5. Semi-elasticities of interval-coded covariates

5.1. Semi-elasticities

In meta-analyses where the user is interested in comparing the semi-elasticities of a certain continuous covariate across different studies, studies that code the covariate of interest in intervals cause a serious problem, as coefficients of interval-coded covariates cannot be compared to coefficients or semi-elasticities of continuous covariates. To overcome this problem, we suggest in this section a procedure to derive a semi-elasticity of an interval-coded covariate.

A regression model with a binary dependent variable, where the kth covariate is interval-coded can be specified as:

$$\Pr(Y=1|X=x) = g\left(\beta_0 + \sum_{j\in\{1,\ldots,K\}\setminus k} \beta_j x_j + \sum_{m\in\{1,\ldots,M\}\setminus m^*} \delta_m D_m\right), \tag{14}$$

$$D_m = \begin{cases} 1 & \text{if } b_{m-1} < x_k \leq b_m \\ 0 & \text{otherwise} \end{cases} \quad \forall\, m = 1,\ldots,M, \tag{15}$$

where $g()$ is a generic link function that can take any form, $x = (x_1,\ldots,x_K)^\top$ is a vector of K continuous covariates, whereas the actual values of one of these covariates, x_k, are unobserved, $D = (D_1,\ldots,D_M)^\top$ is a vector of M dummy variables that indicates in which intervals the values of covariate x_k fall, $b = (b_0,\ldots,b_M)^\top$ is a vector of the $M+1$ boundaries of the M intervals of covariate x_k with $b_0 < b_1 < \ldots < b_{M-1} < b_M$, $m^* \in \{1,\ldots,M\}$ is an arbitrary chosen interval that is used as 'base' interval in the regression, and $\beta = (\beta_0,\ldots,\beta_{k-1},\beta_{k+1},\ldots,\beta_K)^\top$ and $\delta = (\delta_1,\ldots,\delta_{m^*-1},\delta_{m^*+1},\ldots,\delta_M)^\top$ are vectors of K and $M-1$ unknown coefficients, respectively. For convenience of further calculations, we define the (non-estimated) coefficient for the 'base' interval to be zero, i.e., $\delta_{m^*} = 0$.

To derive the semi-elasticity of the 'unknown' continuous kth covariate:

$$\epsilon_k \equiv \frac{\partial \Pr(Y=p|X=x)}{\partial x_k} \cdot x_k, \tag{16}$$

we calculate the effect of an increase of the kth covariate above each inner boundary to the next higher interval on the probability of $Y=1$, i.e.:

$$\begin{aligned} e_{km} = &\Pr(Y=1|b_m < x_k \leq b_{m+1}) - \Pr(Y=1|b_{m-1} < x_k \leq b_m) \\ &\forall\, m = 1,\ldots,M-1 \end{aligned} \tag{17}$$

and the approximate proportions of observations at which the kth covariate will increase above an inner boundary if the kth covariate increases by one percent around each of these boundaries (i.e., the proportions of observations in the intervals $\pm 0.5\%$ around each inner boundary assuming a uniform distribution of the values of the kth covariate within each interval):

$$p_{km} \approx 0.005 \cdot b_m \frac{s_m}{b_m - b_{m-1}} + 0.005 \cdot b_m \frac{s_{m+1}}{b_{m+1} - b_m} \;\forall\, m = 1,\ldots,M-1, \tag{18}$$

where s_m is the proportion of observations that are in the mth interval, i.e., $b_{m-1} <$

$x_k \leq b_m$. Finally, we can calculate the approximate semi-elasticity by:

$$\epsilon_k \approx 100 \cdot \sum_{m=1}^{M-1} e_{km} \cdot p_{km} \tag{19}$$

$$\approx \sum_{m=1}^{M-1} \frac{e_{km} \cdot b_m}{2} \left(\frac{s_m}{b_m - b_{m-1}} + \frac{s_{m+1}}{b_{m+1} - b_m} \right) \tag{20}$$

$$\approx \sum_{m=1}^{M-1} e_{km} w_m \tag{21}$$

$$\text{with } w_m \equiv \frac{b_m}{2} \left(\frac{s_m}{b_m - b_{m-1}} + \frac{s_{m+1}}{b_{m+1} - b_m} \right) \; \forall \, m = 1, \ldots, M \tag{22}$$

so that this semi-elasticity can be interpreted in the same way as the semi-elasticity defined in Section 4, i.e., it indicates the approximate increase in the probability of $Y = 1$ (in percentage points) that is caused by a one percent increase of the covariate x_k.

Table 4 presents the equations for calculating the semi-elasticities of interval-coded covariates for all six estimation methods covered in this article.

To demonstrate how to calculate the semi-elasticity of the interval-coded variable `age` with regard to a married woman's probability to participate in the labour force, we estimate equation (1) as a logistic regression with `age` as interval-coded covariate. We create four intervals, 30–37, 38–44, 45–52, and 53–60 years, and we use the third interval (45–52 years) as 'base' interval in the regression analysis.[9] The results of this estimation are presented in Table 5.

Using the vector of coefficient estimates that can be obtained from Table 5 (`coef(estLogitInt)`) and a vector with the sample means of the covariates `kids` and `educ` and the proportions of observations in the three age intervals included in the regression (`xMeantInt`), one can calculate the semi-elasticity of the covariate `age` using function `urbinElaInt`:

```
urbinElaInt( coef(estLogitInt), xMeanInt, xPos = c( 3, 4, 0, 5 ),
  xBound = c( 30, 37.5, 44.5, 52.5, 60 ), model = "logit" )

##      semEla      stdEr
## -0.3860892          NA
```

Argument `xPos` indicates the positions of the four age intervals (in ascending order) in the vectors `coef(estLogitInt)` and `xMeantInt`, where a zero indicates the position of the reference interval that was not included in the regression and, thus, is not included in `coef(estLogitInt)` or `xMeantInt`. Argument `xBound` indicates the five boundaries of the four intervals. As the coefficients are derived from a logistic regression, we set argument `model` equal to `"logit"`. The semi-elasticity based on the logistic regression with `age` as interval-coded covariate indicates that the probability that a woman is in

[9] The R code for estimating this model is available in Appendix Section C.3.

Table 4: Semi-elasticities of interval-coded covariates for 6 different regression models

Regression model	Semi-elasticity
Linear probability model	$\epsilon_k \approx \sum_{m=1}^{M-1} (\delta_{m+1} - \delta_m)\, w_m$
Binary, bivariate / multivariate, and ordered probit regression	$\epsilon_k \approx \sum_{m=1}^{M-1} (\Phi_{m+1}(\cdot) - \Phi_m(\cdot))\, w_m$ with $\Phi_m(\cdot) \equiv \Phi\left(\beta_0 + \sum_{j \in \{1,\dots,K\} \setminus k} \beta_j x_j + \delta_m\right) \; \forall\, m = 1, \dots, M$
Logistic regression	$\epsilon_k \approx \sum_{m=1}^{M-1} \left(\frac{\exp_{m+1}(\cdot)}{1 + \exp_{m+1}(\cdot)} - \frac{\exp_m(\cdot)}{1 + \exp_m(\cdot)}\right) w_m$ with $\exp_m(\cdot) \equiv \exp\left(\beta_0 + \sum_{j \in \{1,\dots,K\} \setminus k} \beta_j x_j + \delta_m\right) \; \forall\, m = 1, \dots, M$
Multinomial logistic regression	$\epsilon_{k,\mathscr{P}} \approx \sum_{p \in \mathscr{P}} \epsilon^*_{k,p}$ with $\epsilon^*_{k,p} \approx \sum_{m=1}^{M-1} \left(\frac{\exp_{m+1,p}(\cdot)}{\sum_{o=1}^{P} \exp_{m+1,o}(\cdot)} - \frac{\exp_{m,p}(\cdot)}{\sum_{o=1}^{P} \exp_{m,o}(\cdot)}\right) w_m \; \forall\, p = 1, \dots, P$ and $\exp_{m,p}(\cdot) \equiv \exp\left(\beta_{0,p} + \sum_{j \in \{1,\dots,K\} \setminus k} \beta_{j,p} x_j + \delta_{m,p}\right) \; \forall\, m = 1, \dots, M; p = 1, \dots, P$

Note: For multinomial logistic regression models, $\mathscr{P}$ indicates the set of categories of the dependent variable that correspond to a binary outcome of one, while all categories that are not in $\mathscr{P}$ correspond to a binary outcome of zero.

Table 5: Logistic regression results with age as interval-coded covariate

	Dependent variable:
	lfp
Constant	-1.45^{***}
	(0.45)
kids	-0.22^{***}
	(0.06)
age30.37	0.32
	(0.21)
age38.44	0.26
	(0.22)
age53.60	-0.75^{***}
	(0.25)
educ	0.16^{***}
	(0.04)
Observations	753
Log Likelihood	-491.30
Akaike Inf. Crit.	994.59
Note:	*p<0.1; **p<0.05; ***p<0.01

the labour force decreases, *ceteris paribus*, by 0.39 percentage points if her age increases by one percent.

5.2. Approximation of standard errors

An approximate standard error of the semi-elasticity of interval-coded covariates can, again, be obtained by using the Delta method (equation 13). The gradient vectors of the semi-elasticities with respect to the coefficients, $\partial \epsilon_k / \partial (\boldsymbol{\beta}^\top \boldsymbol{\delta}^\top)^\top$, for the various regression models are presented in Appendix Section A.3. Argument `allCoefVcov` of function `urbinElaInt` can be used to specify the variance-covariance matrix:

```
urbinElaInt( coef(estLogitInt), xMeanInt, xPos = c( 3, 4, 0, 5 ),
  xBound = c( 30, 37.5, 44.5, 52.5, 60 ), model = "logit",
  allCoefVcov = vcov(estLogitInt) )

##     semEla      stdEr
## -0.3860892  0.0972512
```

As most studies do not report the variance-covariance matrix, we repeat the above calculation with providing only the standard errors so that `urbinElaInt` sets all covariances to zero:

```
urbinElaInt( coef(estLogitInt), xMeanInt, xPos = c( 3, 4, 0, 5 ),
  xBound = c( 30, 37.5, 44.5, 52.5, 60 ), model = "logit",
  allCoefVcov = sqrt(diag(vcov(estLogitInt))) )

##     semEla      stdEr
## -0.3860892  0.1124600
```

In this empirical example—as in most of our other empirical tests—setting the covariances to zero resulted in a slight overestimation of the standard error and we did not find a way to get better approximations than with just setting the covariances to zero. As setting the covariances to zero usually results only in a slight overestimation of the standard errors, we consider this approximation of the standard errors (which are often anyway only used as weighting factors) as generally suitable for meta-analyses.

6. Effects of continuous covariates when they change between intervals

6.1. Effect size

In this sections, we consider the case where the user wants to compare effects of an interval-coded covariate on the probability of $Y = 1$. We suggest a procedure that uses the results of studies that use the covariate of interest in its continuous form to calculate the effect of this covariate when it switches from one reference interval to another interval.

We start out with a regression equation where the covariate of interest, x_k, is included as a linear term:

$$Pr(Y = 1|X = x) = g\left(\beta_0 + \sum_{k=1}^{K} \beta_k x_k\right) \tag{23}$$

We suggest to derive the (approximate) effects of x_k on Y, if this covariate changes between $M \geq 2$ discrete intervals, e.g., from a 'reference' interval r to an interval of interest l, by:

$$\begin{aligned}
E_{k,lr} = {} & \Pr(Y = 1|b_{l-1} < x_k \leq b_l, x_{-k}) \qquad (24)\\
& - \Pr(Y = 1|b_{r-1} < x_k \leq b_r, x_{-k})\\
= {} & g\left(\beta_0 + \sum_{j\in\{1,\ldots,K\}\setminus k} \beta_j x_j + \beta_k E[x_k|b_{l-1} < x_k \leq b_l]\right) \qquad (25)\\
& - g\left(\beta_0 - \sum_{j\in\{1,\ldots,K\}\setminus k} \beta_j x_j + \beta_k E[x_k|b_{r-1} < x_k \leq b_r]\right)\\
= {} & g\left(\beta_0 + \sum_{j\in\{1,\ldots,K\}\setminus k} \beta_j x_j + \beta_k \bar{x}_{kl}\right) \qquad (26)\\
& - g\left(\beta_0 - \sum_{j\in\{1,\ldots,K\}\setminus k} \beta_j x_j + \beta_k \bar{x}_{kr}\right),
\end{aligned}$$

where $x_{-k} = (x_1, \ldots, x_{k-1}, x_{k+1}, x_K)^\top$ is a vector of all covariates except for x_k, $b_0 < b_1 < \ldots < b_{M-1} < b_M$ are the boundaries of the intervals of covariate x_k, and

$$\bar{x}_{km} \equiv E[x_k|b_{m-1} < x_k \leq b_m] \; \forall \; m = 1, \ldots, M \tag{27}$$

are the expected values of covariate x_k within specific intervals. If the expected values of covariate x_k for specific intervals are unknown, it may be appropriate to approximate them by the mid-points of the respective interval boundaries (e.g., if the covariate x_k has approximately a uniform distribution between the respective interval boundaries):

$$\bar{x}_{km} \approx \frac{b_{m-1} + b_m}{2} \; \forall \; m = 1, \ldots, M. \tag{28}$$

If the model specification additionally includes a quadratic term of the covariate k, e.g., $x_{k+1} = x_k^2$, equations (24) to (26) change to:

$$
\begin{aligned}
E_{k,lr} = \; & \Pr(Y = 1 | b_{l-1} < x_k \leq b_l, x_{-k,k+1}) && (29)\\
& - \Pr(Y = 1 | b_{r-1} < x_k \leq b_r, x_{-k,k+1}) \\
= \; & g\Bigg(\beta_0 + \sum_{j \in \{1,\ldots,K\} \setminus \{k,k+1\}} \beta_j x_j && (30)\\
& + \beta_k E[x_k | b_{l-1} < x_k \leq b_l] + \beta_{k+1} E[x_k^2 | b_{l-1} < x_k \leq b_l] \Bigg) \\
& - g\Bigg(\beta_0 + \sum_{j \in \{1,\ldots,K\} \setminus \{k,k+1\}} \beta_j x_j \\
& + \beta_k E[x_k | b_{r-1} < x_k \leq b_r] + \beta_{k+1} E[x_k^2 | b_{r-1} < x_k \leq b_r] \Bigg) \\
= \; & g\left(\beta_0 + \sum_{j \in \{1,\ldots,K\} \setminus \{k,k+1\}} \beta_j x_j + \beta_k \bar{x}_{kl} + \beta_{k+1} \overline{x_{kl}^2} \right) && (31)\\
& - g\left(\beta_0 + \sum_{j \in \{1,\ldots,K\} \setminus \{k,k+1\}} \beta_j x_j + \beta_k \bar{x}_{kr} + \beta_{k+1} \overline{x_{kr}^2} \right),
\end{aligned}
$$

with

$$\overline{x_{km}^2} \equiv E[x_k^2 | b_{m-1} < x_k \leq b_m] \; \forall \, m = 1, \ldots, M. \tag{32}$$

If $E[x_k^2 | b_{m-1} < x_k \leq b_m]$ is unknown, it may be appropriate to approximate it by assuming that covariate x_k has approximately a uniform distribution between each pair of subsequent interval boundaries so that its probability density function between boundaries b_{m-1} and b_m is $1/(b_m - b_{m-1})$:

$$
\begin{aligned}
\overline{x_{km}^2} & \approx \int_{b_{m-1}}^{b_m} x_k^2 \, \frac{1}{b_m - b_{m-1}} \, d\, x_k && (33)\\
& = \frac{1}{3} \, x_k^3 \, \frac{1}{b_m - b_{m-1}} \Bigg|_{b_{m-1}}^{b_m} && (34)\\
& = \frac{1}{3} \, b_m^3 \, \frac{1}{b_m - b_{m-1}} - \frac{1}{3} \, b_{m-1}^3 \, \frac{1}{b_m - b_{m-1}} && (35)\\
& = \frac{b_m^3 - b_{m-1}^3}{3 \, (b_m - b_{m-1})} \; \forall \, m = 1, \ldots, M. && (36)
\end{aligned}
$$

Table 6 presents the equations for calculating the effect sizes of continuous covariates when they change between intervals for all six estimation methods covered in this article.

The effect of a continuous covariate when it changes between intervals can be calculated with package `urbin` by using function `urbinEffInt`. In our example, we use the results of the probit regression model with variable `age` as a linear covariate as well as

Table 6: Effects of continuous covariates when they change between intervals for 6 different regression models

Regression model	Effect size
Linear probability model	$E_{k,lr} \approx \beta_k \left(\bar{x}_{kl} - \bar{x}_{kr}\right) + \beta_{k+1} \left(\overline{x^2_{kl}} - \overline{x^2_{kr}}\right)$
Binary, bivariate / multivariate, and ordered probit regression	$E_{k,lr} \approx \Phi\left(\beta_0 + \sum_{j\in\{1,\dots,K\}\setminus\{k,k+1\}} \beta_j x_j + \beta_k \bar{x}_{kl} + \beta_{k+1}\overline{x^2_{kl}}\right) - \Phi\left(\beta_0 + \sum_{j\in\{1,\dots,K\}\setminus\{k,k+1\}} \beta_j x_j + \beta_k \bar{x}_{kr} + \beta_{k+1}\overline{x^2_{kr}}\right)$
Logistic regression	$E_{k,lr} \approx \frac{\exp\left(\beta_0 + \sum_{j\in\{1,\dots,K\}\setminus\{k,k+1\}} \beta_j x_j + \beta_k \bar{x}_{kl} + \beta_{k+1}\overline{x^2_{kl}}\right)}{1+\exp\left(\beta_0 + \sum_{j\in\{1,\dots,K\}\setminus\{k,k+1\}} \beta_j x_j + \beta_k \bar{x}_{kl} + \beta_{k+1}\overline{x^2_{kl}}\right)} - \frac{\exp\left(\beta_0 + \sum_{j\in\{1,\dots,K\}\setminus\{k,k+1\}} \beta_j x_j + \beta_k \bar{x}_{kr} + \beta_{k+1}\overline{x^2_{kr}}\right)}{1+\exp\left(\beta_0 + \sum_{j\in\{1,\dots,K\}\setminus\{k,k+1\}} \beta_j x_j + \beta_k \bar{x}_{kr} + \beta_{k+1}\overline{x^2_{kr}}\right)}$
Multinomial logistic regression	$E_{k,lr,\mathscr{P}} \approx \sum_{p\in\mathscr{P}} E^*_{k,lr,p}$ with $E^*_{k,lr,p} \equiv \frac{\exp\left(\beta_{0,p} + \sum_{j\in\{1,\dots,K\}\setminus\{k,k+1\}} \beta_{j,p} x_j + \beta_{k,p} \bar{x}_{kl} + \beta_{k+1,p}\overline{x^2_{kl}}\right)}{\sum_{o=1}^{P}\exp\left(\beta_{0,o} + \sum_{j\in\{1,\dots,K\}\setminus\{k,k+1\}} \beta_{j,o} x_j + \beta_{k,o} \bar{x}_{kl} + \beta_{k+1,o}\overline{x^2_{kl}}\right)} - \frac{\exp\left(\beta_{0,p} + \sum_{j\in\{1,\dots,K\}\setminus\{k,k+1\}} \beta_{j,p} x_j + \beta_{k,p} \bar{x}_{kr} + \beta_{k+1,p}\overline{x^2_{kr}}\right)}{\sum_{o=1}^{P}\exp\left(\beta_{0,o} + \sum_{j\in\{1,\dots,K\}\setminus\{k,k+1\}} \beta_{j,o} x_j + \beta_{k,o} \bar{x}_{kr} + \beta_{k+1,o}\overline{x^2_{kr}}\right)}$

Note: If the regression analysis does not include a quadratic term of the covariate of interest, β_{k+1} is simply the coefficient of the covariate x_{k+1} so that both $\overline{x^2_{kl}}$ and $\overline{x^2_{kr}}$ in the above equations must be replaced by x_{k+1}. For multinomial logistic regression models, $\mathscr{P}$ indicates the set of categories of the dependent variable that correspond to a binary outcome of one, while all categories that are not in $\mathscr{P}$ correspond to a binary outcome of zero.

with `age` as a linear and a quadratic covariate that we already used as example in Section 4 with estimation results presented in Table 3. Based on these estimation results, we calculate the effect of covariate `age` on the probability of women's participation in the labour force as if `age` was an interval-coded covariate and changes from the 30–44 years (reference) interval to the 53–60 years interval. We do this both for the model with `age` as linear covariate:

```
urbinEffInt( coef(estProbit), xMean, xPos = 3,
  refBound = c( 30, 44 ), intBound = c( 53, 60 ),
  model = "probit" )

##     effect      stdEr
## -0.1662336         NA
```

and for the model with `age` as a linear and quadratic covariate:

```
urbinEffInt( coef(estProbitQ), xMeanQ, xPos = c( 3, 4 ),
  refBound = c( 30, 44 ), intBound = c( 53, 60 ),
  model = "probit" )

##     effect      stdEr
## -0.2918354         NA
```

The results based on the two estimated models indicate that the probability that a woman is in the labour force is, *ceteris paribus*, 17 percentage points or 29 percentage points, respectively, lower for women aged 53–60 years than for women aged 30–44 years.

6.2. Approximation of standard errors

As for the semi-elasticities, an approximate standard error of the effect of interval-coded covariates can be obtained by using the Delta method (equation 13). Appendix Section A.4 presents the gradient vectors of the effects with respect to the coefficients, $\partial E_{k,lr} / \partial \boldsymbol{\beta}$, for the various regression models. Argument `allCoefVcov` of function `urbinEffInt` can be used to specify the variance-covariance matrix:

```
urbinEffInt( coef(estProbit), xMean, xPos = 3,
  refBound = c( 30, 44 ), intBound = c( 53, 60 ),
  model = "probit", allCoefVcov = vcov(estProbit) )

##      effect       stdEr
## -0.16623364  0.05243387

urbinEffInt( coef(estProbitQ), xMeanQ, xPos = c( 3, 4 ),
  refBound = c( 30, 44 ), intBound = c( 53, 60 ),
  model = "probit", allCoefVcov = vcov(estProbitQ) )
```

```
##      effect       stdEr
## -0.29183541  0.06370879
```

Given that most studies only report standard errors rather than the (full) variance-covariance matrix, we repeat the above calculations with providing only the standard errors so that `urbinEffInt` sets all covariances to zero:

```
urbinEffInt( coef(estProbit), xMean, xPos = 3,
  refBound = c( 30, 44 ), intBound = c( 53, 60 ),
  model = "probit", allCoefVcov = sqrt(diag(vcov(estProbit))) )

##      effect       stdEr
## -0.16623364  0.05723648

urbinEffInt( coef(estProbitQ), xMeanQ, xPos = c( 3, 4 ),
  refBound = c( 30, 44 ), intBound = c( 53, 60 ),
  model = "probit", allCoefVcov = sqrt(diag(vcov(estProbitQ))) )

## Warning:  In urbinEffInt(allCoef = coef(estProbitQ), allXVal = xMeanQ,
xPos = c(3, 4), refBound = c(30, 44), intBound = c(53, 60), model =
"probit", allCoefVcov = sqrt(diag(vcov(estProbitQ)))) :
##  the returned standard error is likely largely upward biased and,
thus, in most cases meaningless; you can provide the full covariance
matrix via argument 'allCoefVcov' to avoid this bias or use argument
'xMeanSd' to substantially reduce this bias

##     effect      stdEr
## -0.2918354  0.6571220
```

While replacing the (frequently unknown) covariances by zeros usually has only a minor effect on the standard error when the model has only a linear term of the covariate of interest, the standard errors based on models with linear and quadratic terms of the covariate of interest are usually largely upward-biased if the covariances are all set to zeros. However, approximating the covariance between the coefficient of the linear term and the coefficient of the quadratic term as explained in Section 4.2 usually gives sufficiently precise approximations of the standard error. Function `urbinEffInt` applies this procedure, if the user provides the mean value and the standard deviation of the covariate of interest through argument `xMeanSd`:

```
urbinEffInt( coef(estProbitQ), xMeanQ, xPos = c( 3, 4 ),
  refBound = c( 30, 44 ), intBound = c( 53, 60 ),
  model = "probit", allCoefVcov = sqrt(diag(vcov(estProbitQ))),
  xMeanSd = c( mean( Mroz87$age ), sd( Mroz87$age ) ) )

##      effect       stdEr
## -0.29183541  0.07351239
```

7. Grouping and re-basing effects of categorical and interval-coded covariates

7.1. Effect size

In cases where the user is interested in comparing effects of categorical or interval-coded covariates on a binary dependent variable, the user will frequently encounter studies, where the encoding of the covariate of interest differs between studies, e.g., the studies use different reference categories and/or different categorisations.[10] In this section, we suggest an approach to obtain comparable effect sizes by streamlining the categories and unifying the reference category.

We consider a regression model:

$$Pr(Y = 1|X = x) = g\left(\beta_0 + \sum_{j\in\{1,\dots,K\}\setminus k} \beta_j x_j + \sum_{m\in\{1,\dots,M\}\setminus m^*} \delta_m D_m\right), \tag{37}$$

$$D_m = \begin{cases} 1 & \text{if } x_k \in c_m \\ 0 & \text{otherwise} \end{cases} \quad \forall\, m = 1, \dots, M, \tag{38}$$

where $g()$ is again a generic link function and the covariate of interest, x_k, is a categorical variable with M mutually exclusive categories $c_1, \dots, c_M$ with $c_m \cap c_l = \emptyset \;\forall\; m \neq l$, category c_{m^*} is used as reference category, and all other variables and coefficients are defined as above. For notational simplification of the following derivations, we define the (non-estimated) coefficient of the reference category to be zero, i.e., $\delta_{m^*} \equiv 0$.

We want to obtain the effect of a change of covariate x_k from a reference category c_r^* to a category of interest c_l^*:

$$E_{k,lr} = \Pr(Y = 1|x_k \in c_l^*) - \Pr(Y = 1|x_k \in c_r^*), \tag{39}$$

where categories c_r^* and/or c_l^* may comprise multiple original categories $c_1, \dots, c_M$. Vectors $v_r = (v_{r1}, \dots, v_{rM})^\top$ and $v_l = (v_{l1}, \dots, v_{lM})^\top$ indicate, which of the original categories $c_1, \dots, c_M$ are included in categories c_r^* and c_l^*, respectively:

$$v_{nm} = \begin{cases} 1 & \text{if } c_m \in c_n^* \\ 0 & \text{if } c_m \notin c_n^* \end{cases} \forall\, m = 1, \dots, M; n \in \{r, l\} \tag{40}$$

In the following, we derive the effect of a change of covariate x_k from a reference

[10] In order to simplify the notation, we use the term 'categorical variables' throughout this section although all derivations in this section not only apply to (unordered or ordered) categorical variables but equally apply to interval-coded variables.

category c_r^* to a category of interest c_l^*, $E_{k,lr}$ as defined in equation (39):

$$
\begin{aligned}
E_{k,lr} =& \Pr(Y = 1 | x_k \in c_l^*) - \Pr(Y = 1 | x_k \in c_r^*) && (41)\\
=& g\left(\beta_0 + \sum_{j \in \{1,\ldots,K\} \setminus k} \beta_j x_j + \sum_{m=1}^{M} \delta_m E[D_m | x_k \in c_l^*]\right) && (42)\\
& - g\left(\beta_0 + \sum_{j \in \{1,\ldots,K\} \setminus k} \beta_j x_j + \sum_{m=1}^{M} \delta_m E[D_m | x_k \in c_r^*]\right) \\
=& g\left(\beta_0 + \sum_{j \in \{1,\ldots,K\} \setminus k} \beta_j x_j + \sum_{m=1}^{M} \delta_m D_{ml}\right) && (43)\\
& - g\left(\beta_0 + \sum_{j \in \{1,\ldots,K\} \setminus k} \beta_j x_j + \sum_{m=1}^{M} \delta_m D_{mr}\right)
\end{aligned}
$$

with

$$
\begin{aligned}
D_{mn} \equiv& E[D_m | x_k \in c_n^*] && (44)\\
=& \frac{P(D_m = 1 \cap x_k \in c_n^*)}{P(x_k \in c_n^*)} && (45)\\
=& \frac{E[D_m] \, v_{nm}}{P(x_k \in c_n^*)} && (46)\\
=& \frac{s_m v_{nm}}{\sum_{k=1}^{M} s_k v_{nk}} && (47)\\
& \forall \; m = 1, \ldots, M; n \in \{r, l\},
\end{aligned}
$$

where $s_m = E[D_m] \; \forall \; m = 1, \ldots, M$ is the share of observations with covariate x_k being in category c_m.

Table 7 presents the equations for grouping and re-basing effects of categorical and interval-coded covariates for all six estimation methods covered in this article.

To demonstrate how to group and re-base a categorical covariate, we use the results of the logistic regression model with **age** as interval-coded covariate that we already used as example in Section 5 with estimation results presented in Table 5. In this estimation, covariate **age** is coded as four intervals: 30–37 years, 38–44 years, 45–52 years, and 53–60 years, where the interval 45–52 years is used as 'base' interval. In our example, we apply function **urbinEffCat** to group and re-base the categories to calculate the effect of **age** changing from the 30–44 years (reference) interval to the 53–60 years interval:

```
urbinEffCat( coef(estLogitInt), xMeanInt, xPos = c( 3:5 ),
  xGroups = c( -1, -1, 1, 0 ), model = "logit" )

##      effect      stdEr
## -0.2550292          NA
```

Table 7: Grouping and re-basing effects of categorical and interval-coded covariates for 6 different regression models

Regression model	Effect size
Linear probability model	$E_{k,lr} = \sum_{m=1}^{M} \delta_m \left(D_{ml} - D_{mr} \right)$
Binary, bivariate / multivariate, and ordered probit regression	$E_{k,lr} = \Phi \left(\beta_0 + \sum_{j \in \{1,\ldots,K\} \setminus k} \beta_j x_j + \sum_{m=1}^{M} \delta_m D_{ml} \right) - \Phi \left(\beta_0 + \sum_{j \in \{1,\ldots,K\} \setminus k} \beta_j x_j + \sum_{m=1}^{M} \delta_m D_{mr} \right)$
Logistic regression	$E_{k,lr} = \frac{\exp\left(\beta_0 + \sum_{j \in \{1,\ldots,K\} \setminus k} \beta_j x_j + \sum_{m=1}^{M} \delta_m D_{ml}\right)}{1 + \exp\left(\beta_0 + \sum_{j \in \{1,\ldots,K\} \setminus k} \beta_j x_j + \sum_{m=1}^{M} \delta_m D_{ml}\right)} - \frac{\exp\left(\beta_0 + \sum_{j \in \{1,\ldots,K\} \setminus k} \beta_j x_j + \sum_{m=1}^{M} \delta_m D_{mr}\right)}{1 + \exp\left(\beta_0 + \sum_{j \in \{1,\ldots,K\} \setminus k} \beta_j x_j + \sum_{m=1}^{M} \delta_m D_{mr}\right)}$
Multinomial logistic regression	$E_{k,lr,\mathscr{P}} = \sum_{p \in \mathscr{P}} E^*_{k,lr,p}$ $E^*_{k,lr,p} = \frac{\exp\left(\beta_{0,p} + \sum_{j \in \{1,\ldots,K\} \setminus k} \beta_{j,p} x_j + \sum_{m=1}^{M} \delta_{m,p} D_{ml}\right)}{\sum_{o=1}^{P} \exp\left(\beta_{0,o} + \sum_{j \in \{1,\ldots,K\} \setminus k} \beta_{j,o} x_j + \sum_{m=1}^{M} \delta_{m,o} D_{ml}\right)} - \frac{\exp\left(\beta_{0,p} + \sum_{j \in \{1,\ldots,K\} \setminus k} \beta_{j,p} x_j + \sum_{m=1}^{M} \delta_{m,p} D_{mr}\right)}{\sum_{o=1}^{P} \exp\left(\beta_{0,o} + \sum_{j \in \{1,\ldots,K\} \setminus k} \beta_{j,o} x_j + \sum_{m=1}^{M} \delta_{m,o} D_{mr}\right)}$

Note: For multinomial logistic regression models, $\mathscr{P}$ indicates the set of categories of the dependent variable that correspond to a binary outcome of one, while all categories that are not in $\mathscr{P}$ correspond to a binary outcome of zero.

Argument `xPos` indicates the positions of the categories of variable `age` in the coefficient vector and in the vector of mean values, argument `xGroups` indicates how the four original categories should be grouped and re-based, and all other arguments are defined as explained for the other functions of the `urbin` package. Argument `xGroups` must have one element for each category that was used in the estimation, where the categories are in the same order as indicated by argument `xPos` and the last element is the 'base' category, i.e., in our case, the elements of argument `xGroups` must correspond to the order: 30–37 years (first element of `xPos`), 38–44 years (second element of `xPos`), 53–60 years (third element of `xPos`), and 45–52 years (reference category). Each element of argument `xGroups` must be a -1 (indicating that the category should belong to the new reference category), a 1 (indicating that the category should belong to the new category of interest), or a 0 (indicating that the category should neither belong to the new reference category nor to the new category of interest). As the new reference category comprises both the 30–37 years interval and the 38–44 years interval, the values of the first two elements of argument `xGroups` must be -1. As the new category of interest is the 53–60 years interval, the value of the third element of argument `xGroups` must be 1. As the old reference interval, 45–52 years, is neither in the new reference category nor in the new category of interest, the value of the fourth element of argument `xGroups` must be 0. The calculated effect size indicates that the probability that a woman is in the labour force is, *ceteris paribus*, 26 percentage points lower for women aged 53–60 years than for women aged 30–44 years.

7.2. Approximation of standard errors

An approximate standard error of the effect of a grouped and re-based covariate can, again, be obtained by using the Delta method (equation 13). Appendix Section A.5 presents the gradient vectors of the effects with respect to the coefficients, $\partial E_{k,lr} / \partial(\boldsymbol{\beta}^\top \boldsymbol{\delta}^\top)^\top$, for the various regression models. Argument `allCoefVcov` of function `urbinEffCat` can be used to specify the variance-covariance matrix of the estimated coefficients:

```
urbinEffCat( coef(estLogitInt), xMeanInt, c( 3:5 ),
  c( -1, -1, 1, 0 ), vcov(estLogitInt), model = "logit" )

##      effect       stdEr
## -0.25502923  0.06231656
```

As most studies do not report the variance-covariance matrix, we repeat the previous calculation with providing only the standard errors so that `urbinEffCat` sets all covariances to zero:

```
urbinEffCat( coef(estLogitInt), xMeanInt, c( 3:5 ),
  c( -1, -1, 1, 0 ), sqrt(diag(vcov(estLogitInt))),
  model = "logit" )
```

```
##      effect       stdEr
## -0.25502923  0.06958641
```

Similarly to Section 5, setting the covariances to zero usually results in a slight overestimation of the standard errors. As these overestimations are usually small and the standard errors are often anyway only used as weighting factors, we consider this approximation of the standard errors to be generally suitable for meta-analyses.

8. Non-binary categorical dependent variables

As explained in Section 3, it can be possible to make results of studies with non-binary categorical dependent variables comparable to results of studies with binary dependent variables, if the categories of the (non-binary) dependent variable can be grouped into two groups that correspond to the two outcomes of the binary dependent variable in the other studies.

In order to demonstrate this, we use an ordered probit regression with `age` as linear and quadratic covariate and a multinomial logistic regression with `age` as interval-coded covariate as examples.[11] The estimation results of these two models are presented in Tables 8 and 9, respectively.

We combine the two outcomes 'part-time labour force participation' and 'full-time labour force participation' to one joint outcome category so that we obtain a binary outcome: 'no labour force participation' and (part-time or full-time) 'labour force participation'. For ordered probit models, the negative value of the break point that separates the two groups of categories corresponds to the intercept of a binary probit model (see Section 3). Hence, in our example, the relevant break point is the one between the 'no labour force participation' category and the 'part-time labour force participation' category, which has an estimated value of 3.13 (see Table 8). When applying one of the functions of the `urbin` package to ordered probit models, argument `iPos` must indicate the position of this break point in the vector of coefficients, while all other break points must be ignored. The element in the vector of the values of the covariates that corresponds to the relevant break point (as indicated by argument `iPos`) must be minus one, in order to take into account that the intercept of a corresponding binary probit model must be replaced by the *negative* value of the relevant break point of an ordered probit model. We set argument `model` to `"oprobit"` to indicate an ordered probit model, while all other arguments are used as explained above:

```
urbinEla( coef(summary(estOProbitQ))[-6,1], c( xMeanQ[-1], -1 ),
  xPos = c( 2, 3 ), iPos = 5, model = "oprobit",
  vcov(estOProbitQ)[-6,-6] )

##     semEla      stdEr
## -0.3467696  0.1201219
```

[11] The R code for estimating these two models is available in Appendix Sections C.4 and C.5, respectively.

Table 8: Ordered probit regression results with age as linear and quadratic covariate

	Dependent variable:
	lfp3
kids	-0.18^{***}
	(0.04)
age	0.16^{***}
	(0.02)
I(ageˆ2)	-0.002^{***}
	(0.0003)
educ	0.07^{***}
	(0.02)
no|part	3.13^{***}
	(0.001)
part|full	3.87^{***}
	(0.05)
Observations	753
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 9: Multinomial logistic regression results with age as interval-coded covariate

	Dependent variable:
	lfp3
(Intercept):full	-1.73^{***}
	(0.53)
(Intercept):part	-2.62^{***}
	(0.55)
kids:full	-0.42^{***}
	(0.08)
kids:part	-0.06
	(0.07)
age30.37TRUE:full	0.57^{**}
	(0.25)
age30.37TRUE:part	0.09
	(0.25)
age38.44TRUE:full	0.49^{*}
	(0.26)
age38.44TRUE:part	0.04
	(0.27)
age53.60TRUE:full	-0.75^{**}
	(0.30)
age53.60TRUE:part	-0.76^{**}
	(0.33)
educ:full	0.15^{***}
	(0.04)
educ:part	0.19^{***}
	(0.04)
Observations	753
R^2	0.04
Log Likelihood	-778.06
LR Test	66.03^{***} (df = 12)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

The calculated semi-elasticity indicates that the probability that a woman is at least part-time in the labour force decreases, *ceteris paribus*, by 0.35 percentage points if her age increases by one percent.

In the multivariate logistic regression, 'no labour force participation' is used as the reference category of the dependent variables, while 'full-time labour force participation' and 'part-time labour force participation' are used as first alternative category and second alternative category, respectively (see Table 9). When applying one of the functions of the urbin package to a multinomial logistic regression, argument yCat must indicate the categories of the dependent variable $\mathscr{P}$ that correspond to a binary outcome of one. All other categories are considered to correspond to a binary outcome of zero. In argument yCat, a zero indicates the reference category, while a one, two, or three, etc. indicates the first, second, or third, etc. alternative category, respectively. As the first and second alternative categories comprise the binary outcome of one in our example, i.e., $\mathscr{P} = \{1, 2\}$, argument yCat must be a vector with two values: one and two. We set argument model to "mlogit" to indicate a multinomial logistic regression, while all other arguments are used as explained above:

```
coefPermuteInt <- c( seq( 1, 11, 2 ), seq( 2, 12, 2 ) )
urbinElaInt( coef(estMLogitInt)[coefPermuteInt], xMeanInt,
  c( 3, 4, 0, 5 ), c( 30, 37.5, 44.5, 52.5, 60 ), model = "mlogit",
  vcov(estMLogitInt)[coefPermuteInt,coefPermuteInt],
  yCat = c( 1, 2 ) )

##      semEla       stdEr
## -0.39395280  0.09774856
```

As the functions in package urbin expect that the coefficients of multinomial logistic regressions are grouped for each category of the dependent variable (i.e., $\beta_{0,1}, \ldots, \beta_{K,1}, \beta_{0,2}, \ldots, \beta_{K,2}, \ldots, \beta_{0,P}, \ldots, \beta_{K,P}$), while the coefficients of models estimated by mlogit are grouped for each covariate (i.e., $\beta_{0,1}, \ldots, \beta_{0,P}, \beta_{1,1}, \ldots, \beta_{1,P}, \ldots, \beta_{K,1}, \ldots, \beta_{K,P}$), we created a vector coefPermuteInt that reorders the coefficients and their variances and covariances so that they are ordered as expected by package urbin. The semi-elasticity indicates that the probability that a woman is either part-time or full-time in the labour force decreases, *ceteris paribus*, by 0.39 percentage points if her age increases by one percent.

9. Conclusion

The direct comparison of coefficients from regression analyses from different studies is often meaningless because the studies use different estimation methods or different units of measurements or different encodings of the variables of interest. In this article, we propose straightforward and easy-to-implement approaches to unify results from regression analyses with binary dependent variables or categorical dependent variables that can be transformed to binary variables.

We have implemented all suggested approaches in the R package urbin. This article uses this package to demonstrate how regression results from differently specified regression analyses can be unified by calculating semi-elasticities of continuous and interval-coded covariates, by calculating effects of continuous covariates when they change between intervals, and by grouping and re-basing effects of categorical and interval-coded covariates. We show how to obtain valid approximations for the calculated standard errors of the semi-elasticities and effect sizes without information about the variance-covariance matrix of the coefficients, e.g., for cases where the user wants to use the standard errors as weighting factors in a meta-analysis.

References

[1] Mroz, TA. The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica*, 1987; 55: 765–799.

[2] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2018.

[3] Henningsen, A, Henningsen, G. *urbin: Unifying Estimation Results with Binary Dependent Variables.* R package version 0.1-6, available at `https://CRAN.R-project.org/package=urbin`.

[4] Toomet, O, Henningsen, A. Sample selection models in R: Package sampleSelection. *Journal of Statistical Software*, 2008; 27(7).

[5] Greene, WH. *Econometric Analysis.* 7th ed. Pearson, 2011.

[6] Bliss, CI. The method of probits. *Science*, 1934; 79: 38–39.

[7] Fisher, RA. The case of zero survivors in probit assays. *Annals of Applied Biology*, 1935; 22: 164–165.

[8] Cox, D. The regression analysis of binary sequences (with discussion). *Journal of the Royal Statistical Society*, 1958; 20(2): 215–242.

[9] Sodjinou, E, Henningsen, A. Community-based management and interrelations between different technology adoption decisions: Innovations in village poultry farming in western africa. Tech. Rep. 2012/11, IFRO Working Paper, 2012.

[10] Aitchison, J, Silvey, DS. The generalization of probit analysis to the case of multiple repsonses. *Biometrika*, 1957; 44: 131–140.

[11] Luce, RD. *Individual Choice Behavior: A Theoretical Analysis.* New York: John Wiley, 1959.

[12] Greene, WH, Hensher, DA. *Modeling Ordered Choices: A Primer.* Cambridge University Press, 2010.

[13] Train, K. *Discrete Choice Methods with Simulation.* Cambridge University Press, 2002.

A. Gradients for calculating approximate standard errors

A.1. Gradients of semi-elasticities of continuous covariates

A.1.1. Linear probability model

If the regression equation includes only a linear term of the covariate of interest, the gradients are:

$$\frac{\partial \epsilon_k}{\partial \beta_j} = 0 \; \forall \; j \in \{0, \ldots, K\} \setminus k \tag{48}$$

$$\frac{\partial \epsilon_k}{\partial \beta_k} = x_k. \tag{49}$$

If the regression equation additionally includes a quadratic term of the covariate of interest, there is one additional gradient:

$$\frac{\partial \epsilon_k}{\partial \beta_{k+1}} = 2 \; x_k^2. \tag{50}$$

A.1.2. Probit regression

If the regression equation includes only a linear term of the covariate of interest, the gradients are:

$$\frac{\partial \epsilon_k}{\partial \beta_j} = -\boldsymbol{X}'\boldsymbol{\beta} \; \epsilon_k \; x_j \; \forall \; j \in \{0, \ldots, K\} \setminus k \tag{51}$$

$$\frac{\partial \epsilon_k}{\partial \beta_k} = -\boldsymbol{X}'\boldsymbol{\beta} \; \epsilon_k \; x_k + \phi(\boldsymbol{X}'\boldsymbol{\beta}) \; x_k \tag{52}$$

with $x_0 \equiv 1$.

If the regression equation additionally includes a quadratic term of the covariate of interest, there is one additional gradient:

$$\frac{\partial \epsilon_k}{\partial \beta_{k+1}} = -\boldsymbol{X}'\boldsymbol{\beta} \; \epsilon_k \; x_k^2 + 2 \; \phi(\boldsymbol{X}'\boldsymbol{\beta}) \; x_k^2. \tag{53}$$

A.1.3. Logistic regression

If the regression equation includes only a linear term of the covariate of interest, the gradients are:

$$\frac{\partial \epsilon_k}{\partial \beta_j} = \left(1 - \frac{2 \; \exp\left(\boldsymbol{X}'\boldsymbol{\beta}\right)}{1 + \exp\left(\boldsymbol{X}'\boldsymbol{\beta}\right)} \right) \epsilon_k \; x_j \; \forall \; j \in \{0, \ldots, K\} \setminus k \tag{54}$$

$$\frac{\partial \epsilon_k}{\partial \beta_k} = \left(1 - \frac{2 \; \exp\left(\boldsymbol{X}'\boldsymbol{\beta}\right)}{1 + \exp\left(\boldsymbol{X}'\boldsymbol{\beta}\right)} \right) \epsilon_k \; x_k + \frac{\exp\left(\boldsymbol{X}'\boldsymbol{\beta}\right)}{\left(1 + \exp\left(\boldsymbol{X}'\boldsymbol{\beta}\right)\right)^2} \; x_k \tag{55}$$

with $x_0 \equiv 1$ (see also [13]).

If the regression equation additionally includes a quadratic term of the covariate of interest, there is one additional gradient:

$$\frac{\partial \epsilon_k}{\partial \beta_{k+1}} = \left(1 - \frac{2\ \exp\left(\boldsymbol{X}'\boldsymbol{\beta}\right)}{1 + \exp\left(\boldsymbol{X}'\boldsymbol{\beta}\right)}\right) \epsilon_k\ x_k^2 + 2\ \frac{\exp\left(\boldsymbol{X}'\boldsymbol{\beta}\right)}{\left(1 + \exp\left(\boldsymbol{X}'\boldsymbol{\beta}\right)\right)^2} x_k^2. \tag{56}$$

A.1.4. Multinomial logistic regression

If the regression equation includes only a linear term of the covariate of interest, the gradients are:

$$\frac{\partial \epsilon_{k,\mathscr{P}}}{\partial \beta_{j,o}} = \sum_{p \in \mathscr{P}} \frac{\partial \epsilon^*_{k,p}}{\partial \beta_{j,o}}\ \forall\ j = 0, \ldots, K; o \in \{1, \ldots, P\} \setminus p^* \tag{57}$$

$$\text{with } \frac{\partial \epsilon^*_{k,p}}{\partial \beta_{j,o}} = \left(-\pi_p \epsilon^*_{k,o} - \pi_o \epsilon^*_{k,p} + \Delta_{o,p} \epsilon^*_{k,p}\right) x_j \tag{58}$$

$$\forall\ j \in \{0, \ldots, K\} \setminus k; p = 1, \ldots, P; o \in \{1, \ldots, P\} \setminus p^*,$$

$$\frac{\partial \epsilon^*_{k,p}}{\partial \beta_{k,o}} = \left(-\pi_p \epsilon^*_{k,o} - \pi_o \epsilon^*_{k,p} - \pi_p \pi_o + \Delta_{o,p} \left(\pi_p + \epsilon^*_{k,p}\right)\right) x_k \tag{59}$$

$$\forall\ p = 1, \ldots, P; o \in \{1, \ldots, P\} \setminus p^*,$$

$x_0 \equiv 1$, and $\Delta_{o,p}$ denoting Kronecker's Delta with $\Delta_{o,p} = 1\ \forall\ o = p$ and $\Delta_{o,p} = 0\ \forall\ o \neq p$.

If the regression equation additionally includes a quadratic term of the covariate of interest, there are $P - 1$ additional gradients:

$$\frac{\partial \epsilon_{k,\mathscr{P}}}{\partial \beta_{k+1,o}} = \sum_{p \in \mathscr{P}} \frac{\partial \epsilon^*_{k,p}}{\partial \beta_{k+1,o}}\ \forall\ o \in \{1, \ldots, P\} \setminus p^* \tag{60}$$

$$\text{with } \frac{\partial \epsilon^*_{k,p}}{\partial \beta_{k+1,o}} = \left(-\pi_p \epsilon^*_{k,o} - \pi_o \epsilon^*_{k,p} - 2\pi_p \pi_o + \Delta_{o,p} \left(2\pi_p + \epsilon^*_{k,p}\right)\right) x_k^2 \tag{61}$$

$$\forall\ p = 1, \ldots, P; o \in \{1, \ldots, P\} \setminus p^*.$$

A.2. Simplified gradients of semi-elasticities of continuous covariates

A.2.1. Linear probability model

As almost all elements of the gradient vector are zero (see section A.1.1), almost all off-diagonal elements of the variance-covariance matrix of the estimated coefficients are anyway ignored when the Delta method is applied to calculate the approximate standard error of the semi-elasticity. Therefore, we do not need to obtain 'simplified' gradients in order to avoid biases due to missing information about the off-diagonal elements of the variance-covariance matrix.

A.2.2. Probit regression

In order to improve the approximation of the standard errors when the off-diagonal elements of the variance-covariance matrix of the estimated coefficients are unknown, we simplify the derivation of the gradients by ignoring that the 'weighting factor' $\phi(\cdot)$ in the equation for calculating the semi-elasticities (see Table 2) depends on the coefficients. If the regression equation includes only a linear term of the covariate of interest, the 'simplified' gradients are:

$$\frac{\partial \epsilon_k}{\partial \beta_j} = 0 \; \forall \; j \in \{0, \ldots, K\} \setminus k \tag{62}$$

$$\frac{\partial \epsilon_k}{\partial \beta_k} = \phi(\boldsymbol{X}'\boldsymbol{\beta}) \; x_k. \tag{63}$$

If the regression equation additionally includes a quadratic term of the covariate of interest, there is one additional 'simplified' gradient:

$$\frac{\partial \epsilon_k}{\partial \beta_{k+1}} = 2 \; \phi(\boldsymbol{X}'\boldsymbol{\beta}) \; x_k^2. \tag{64}$$

A.2.3. Logistic regression

In order to improve the approximation of the standard errors when the off-diagonal elements of the variance-covariance matrix of the estimated coefficients are unknown, we simplify the derivation of the gradients by ignoring that the 'weighting factor' $\exp(\cdot)/(1-\exp(\cdot))^2$ in the equation for calculating the semi-elasticities (see Table 2) depends on the coefficients. If the regression equation includes only a linear term of the covariate of interest, the 'simplified' gradients are:

$$\frac{\partial \epsilon_k}{\partial \beta_j} = 0 \; \forall \; j \in \{0, \ldots, K\} \setminus k \tag{65}$$

$$\frac{\partial \epsilon_k}{\partial \beta_k} = \frac{\exp\left(\boldsymbol{X}'\boldsymbol{\beta}\right)}{\left(1 + \exp\left(\boldsymbol{X}'\boldsymbol{\beta}\right)\right)^2} \; x_k \tag{66}$$

If the regression equation additionally includes a quadratic term of the covariate of interest, there is one additional 'simplified' gradient:

$$\frac{\partial \epsilon_k}{\partial \beta_{k+1}} = 2 \; \frac{\exp\left(\boldsymbol{X}'\boldsymbol{\beta}\right)}{\left(1 + \exp\left(\boldsymbol{X}'\boldsymbol{\beta}\right)\right)^2} x_k^2. \tag{67}$$

A.2.4. Multinomial logistic regression

In order to improve the approximation of the standard errors when the off-diagonal elements of the variance-covariance matrix of the estimated coefficients are unknown, we simplify the derivation of the gradients by ignoring that the 'weighting factors' $\pi_p; p = \{1, \ldots, P\}$ and $\pi_o; o = \{1, \ldots, P\}$ in the equation for calculating the semi-elasticities (see Table 2) depend on the coefficients. If the regression equation includes only a linear term of the covariate of interest, the 'simplified' gradients are:

$$\frac{\partial \epsilon_{k,\mathscr{P}}}{\partial \beta_{j,o}} = \sum_{p \in \mathscr{P}} \frac{\partial \epsilon^*_{k,p}}{\partial \beta_{j,o}} \; \forall \; j = 0, \ldots, K; o \in \{1, \ldots, P\} \setminus p^* \tag{68}$$

$$\text{with } \frac{\partial \epsilon^*_{k,p}}{\partial \beta_{j,o}} = 0 \; \forall \; j \in \{0, \ldots, K\} \setminus k; p = 1, \ldots, P; o \in \{1, \ldots, P\} \setminus p^*, \tag{69}$$

$$\frac{\partial \epsilon^*_{k,p}}{\partial \beta_{k,o}} = \left(-\pi_p \pi_o + \Delta_{o,p} \pi_p\right) x_k \; \forall \; p = 1, \ldots, P; o \in \{1, \ldots, P\} \setminus p^*, \tag{70}$$

and $\Delta_{o,p}$ denoting Kronecker's Delta with $\Delta_{o,p} = 1 \; \forall \; o = p$ and $\Delta_{o,p} = 0 \; \forall \; o \neq p$.

If the regression equation additionally includes a quadratic term of the covariate of interest, there are $P - 1$ additional gradients:

$$\frac{\partial \epsilon_{k,\mathscr{P}}}{\partial \beta_{k+1,o}} = \sum_{p \in \mathscr{P}} \frac{\partial \epsilon^*_{k,p}}{\partial \beta_{k+1,o}} \; \forall \; o \in \{1, \ldots, P\} \setminus p^* \tag{71}$$

$$\text{with } \frac{\partial \epsilon^*_{k,p}}{\partial \beta_{k+1,o}} = \left(-2\pi_p \pi_o + 2\Delta_{o,p} \pi_p\right) x_k^2 \quad \forall \; p = 1, \ldots, P; o \in \{1, \ldots, P\} \setminus p^*. \tag{72}$$

A.3. Gradients of semi-elasticities of interval-coded covariates

A.3.1. Linear probability model

The gradients are:

$$\frac{\partial \epsilon_k}{\partial \beta_j} = 0 \; \forall \; j \in \{0, \ldots, K\} \setminus k \tag{73}$$

$$\frac{\partial \epsilon_k}{\partial \delta_1} = -w_1 \tag{74}$$

$$\frac{\partial \epsilon_k}{\partial \delta_m} = w_{m-1} - w_m \; \forall \; m \in \{2, \ldots, M-1\} \setminus m^* \tag{75}$$

$$\frac{\partial \epsilon_k}{\partial \delta_M} = w_{M-1} \tag{76}$$

A.3.2. Probit regression

The gradients are:

$$\frac{\partial \epsilon_k}{\partial \beta_j} = x_j \sum_{m=1}^{M-1} \left(\phi_{m+1}(\cdot) - \phi_m(\cdot)\right) w_m \;\forall\, j \in \{0, \ldots, K\} \setminus k \tag{77}$$

$$\frac{\partial \epsilon_k}{\partial \delta_1} = \; - \phi_1(\cdot)\; w_1 \tag{78}$$

$$\frac{\partial \epsilon_k}{\partial \delta_m} = \phi_m(\cdot)\,(w_{m-1} - w_m)\;\; \forall\, m \in \{2, \ldots, M-1\} \setminus m^* \tag{79}$$

$$\frac{\partial \epsilon_k}{\partial \delta_M} = \phi_M(\cdot)\; w_{M-1} \tag{80}$$

$$\text{with } \phi_m \equiv \phi\left(\beta_0 + \sum_{j \in \{1,\ldots,K\}\setminus k} \beta_j x_j + \delta_m\right)\;\; \forall\, m = 1, \ldots, M \tag{81}$$

and $x_0 \equiv 1$.

A.3.3. Logistic regression

The gradients are:

$$\frac{\partial \epsilon_k}{\partial \beta_j} = x_j \sum_{m=1}^{M-1} \left(\frac{\exp_{m+1}(\cdot)}{(1+\exp_{m+1}(\cdot))^2} - \frac{\exp_m(\cdot)}{(1+\exp_m(\cdot))^2}\right) w_m \tag{82}$$

$$\forall\, j \in \{0, \ldots, K\} \setminus k$$

$$\frac{\partial \epsilon_k}{\partial \delta_1} = \; - \frac{\exp_1(\cdot)}{(1+\exp_1(\cdot))^2}\; w_1 \tag{83}$$

$$\frac{\partial \epsilon_k}{\partial \delta_m} = \frac{\exp_m(\cdot)}{(1+\exp_m(\cdot))^2}\,(w_{m-1} - w_m)\;\; \forall\, m \in \{2, \ldots, M-1\} \setminus m^* \tag{84}$$

$$\frac{\partial \epsilon_k}{\partial \delta_M} = \frac{\exp_M(\cdot)}{(1+\exp_M(\cdot))^2}\; w_{M-1} \tag{85}$$

with $x_0 \equiv 1$ and $\exp_m; m = 1, \ldots, M$ as defined in Table 4.

A.3.4. Multinomial logistic regression

The gradients are:

$$\frac{\partial \epsilon_{k,\mathscr{P}}}{\partial \beta_{j,o}} = \sum_{p\in\mathscr{P}} \frac{\partial \epsilon^*_{k,p}}{\partial \beta_{j,o}} \;\forall\, j \in \{0,\ldots,K\} \setminus k; o \in \{1,\ldots,P\} \setminus p^* \tag{86}$$

$$\frac{\partial \epsilon_{k,\mathscr{P}}}{\partial \delta_{m,o}} = \sum_{p\in\mathscr{P}} \frac{\partial \epsilon^*_{k,p}}{\partial \delta_{m,o}} \;\forall\, m \in \{1,\ldots,M\} \setminus m^*; o \in \{1,\ldots,P\} \setminus p^* \tag{87}$$

$$\text{with } \frac{\partial \epsilon^*_{k,p}}{\partial \beta_{j,o}} = x_j \sum_{m=1}^{M-1} \Big(\pi_{p,m}\pi_{o,m} - \pi_{p,m+1}\pi_{o,m+1} - \Delta_{op}\left(\pi_{p,m} - \pi_{p,m+1}\right)\Big) w_m \tag{88}$$

$$\forall\, j \in \{0,\ldots,K\} \setminus k; p = 1,\ldots,P; o \in \{1,\ldots,P\} \setminus p^*,$$

$$\frac{\partial \epsilon^*_{k,p}}{\partial \delta_{1,o}} = \left(\pi_{p,1}\pi_{o,1} - \Delta_{o,p}\,\pi_{p,1}\right) w_1 \tag{89}$$

$$\forall\, p = 1,\ldots,P; o \in \{1,\ldots,P\} \setminus p^*,$$

$$\frac{\partial \epsilon^*_{k,p}}{\partial \delta_{m,o}} = \left(\pi_{p,m}\pi_{o,m} - \Delta_{o,p}\,\pi_{p,m}\right)\left(w_m - w_{m-1}\right) \tag{90}$$

$$\forall\, m \in \{2,\ldots,M-1\} \setminus m^*; p = 1,\ldots,P; o \in \{1,\ldots,P\} \setminus p^*,$$

$$\frac{\partial \epsilon^*_{k,p}}{\partial \delta_{M,o}} = -\left(\pi_{p,M}\pi_{o,M} - \Delta_{o,p}\,\pi_{p,M}\right) w_{M-1} \tag{91}$$

$$\forall\, p = 1,\ldots,P; o \in \{1,\ldots,P\} \setminus p^*,$$

$$\pi_{p,m} \equiv \frac{\exp_{m,p}}{\sum_{o=1}^{P} \exp_{m,o}} \;\forall\, p = 1,\ldots,P; m = 1,\ldots,M, \tag{92}$$

$x_0 \equiv 1$, and $\exp_{m,p}; m = 1,\ldots,M; p = 1,\ldots,P$ as defined in Table 4.

A.4. Gradients of effects of continuous covariates when they change between intervals

A.4.1. Linear probability model

If the regression equation includes only a linear term of the covariate of interest, the gradients are:

$$\frac{\partial E_{k,lr}}{\partial \beta_j} = 0 \;\forall\, j \in \{0,\ldots,K\} \setminus k \tag{93}$$

$$\frac{\partial E_{k,lr}}{\partial \beta_k} = \bar{x}_{kl} - \bar{x}_{kr}. \tag{94}$$

If the regression equation additionally includes a quadratic term of the covariate of interest, there is one additional gradient:

$$\frac{\partial E_{k,lr}}{\partial \beta_{k+1}} = \overline{x^2_{kl}} - \overline{x^2_{kr}}. \tag{95}$$

A.4.2. Probit regression

If the regression equation includes only a linear term of the covariate of interest, the gradients are:

$$\frac{\partial E_{k,lr}}{\partial \beta_j} = (\phi_l - \phi_r)\, x_j \ \forall\, j \in \{0, \ldots, K\} \setminus k \tag{96}$$

$$\frac{\partial E_{k,lr}}{\partial \beta_k} = \phi_l\, \bar{x}_{kl} - \phi_r\, \bar{x}_{kr} \tag{97}$$

$$\text{with } \phi_n \equiv \phi\left(\beta_0 + \sum_{j \in \{1,\ldots,K\}\setminus k} \beta_j x_j + \beta_k \bar{x}_{kn}\right) \ \forall\, n \in \{l, r\} \tag{98}$$

and $x_0 \equiv 1$.

If the regression equation additionally includes a quadratic term of the covariate of interest, we have:

$$\phi_n \equiv \phi\left(\beta_0 + \sum_{j \in \{1,\ldots,K\}\setminus\{k,k+1\}} \beta_j x_j + \beta_k \bar{x}_{kn} + \beta_{k+1}\overline{x^2}_{kn}\right) \quad \forall\, n \in \{l, r\} \tag{99}$$

and there is one additional gradient:

$$\frac{\partial E_{k,lr}}{\partial \beta_{k+1}} = \phi_l\, \overline{x^2}_{kl} - \phi_r\, \overline{x^2}_{kr}. \tag{100}$$

A.4.3. Logistic regression

If the regression equation includes only a linear term of the covariate of interest, the gradients are:

$$\frac{\partial E_{k,lr}}{\partial \beta_j} = \left(\frac{\exp_l(\cdot)}{(1+\exp_l(\cdot))^2} - \frac{\exp_r(\cdot)}{(1+\exp_r(\cdot))^2}\right) x_j \ \forall\, j \in \{0, \ldots, K\} \setminus k \tag{101}$$

$$\frac{\partial E_{k,lr}}{\partial \beta_k} = \frac{\exp_l(\cdot)}{(1+\exp_l(\cdot))^2}\, \bar{x}_{kl} - \frac{\exp_r(\cdot)}{(1+\exp_r(\cdot))^2}\, \bar{x}_{kr} \tag{102}$$

$$\text{with } \exp_n \equiv \exp\left(\beta_0 + \sum_{j \in \{1,\ldots,K\}\setminus k} \beta_j x_j + \beta_k \bar{x}_{kn}\right) \ \forall\, n \in \{l, r\} \tag{103}$$

and $x_0 \equiv 1$.

If the regression equation additionally includes a quadratic term of the covariate of interest, we have:

$$\exp_n \equiv \exp\left(\beta_0 + \sum_{j\in\{1,\ldots,K\}\setminus\{k,k+1\}} \beta_j x_j + \beta_k \bar{x}_{kn} + \beta_{k+1}\overline{x^2}_{kn}\right) \quad \forall\, n \in \{l, r\} \tag{104}$$

and there is one additional gradient:

$$\frac{\partial E_{k,lr}}{\partial \beta_{k+1}} = \frac{\exp_l(\cdot)}{(1+\exp_l(\cdot))^2}\,\overline{x^2}_{kl} - \frac{\exp_r(\cdot)}{(1+\exp_r(\cdot))^2}\,\overline{x^2}_{kr}. \tag{105}$$

A.4.4. Multinomial logistic regression

If the regression equation includes only a linear term of the covariate of interest, the gradients are:

$$\frac{\partial E_{k,lr,\mathscr{P}}}{\partial \beta_{j,o}} = \sum_{p\in\mathscr{P}} \frac{\partial E^*_{k,lr,p}}{\partial \beta_{j,o}} \;\forall\, j = 0,\ldots,K; o \in \{1,\ldots,P\}\setminus p^* \tag{106}$$

$$\frac{\partial E^*_{k,lr,p}}{\partial \beta_{j,o}} = \Big(\pi_{p,r}\,\pi_{o,r} - \pi_{p,l}\,\pi_{o,l} - \Delta_{o,p}\,(\pi_{p,r} - \pi_{p,l})\Big)x_j \quad \forall\, j \in \{0,\ldots,K\}\setminus k; p = 1,\ldots,P; o \in \{1,\ldots,P\}\setminus p^* \tag{107}$$

$$\frac{\partial E^*_{k,lr,p}}{\partial \beta_{k,o}} = (\pi_{p,r}\,\pi_{o,r} - \Delta_{o,p}\,\pi_{p,r})\,\bar{x}_{kr} - (\pi_{p,l}\,\pi_{o,l} - \Delta_{o,p}\,\pi_{p,l})\,\bar{x}_{kl} \quad \forall\, p = 1,\ldots,P; o \in \{1,\ldots,P\}\setminus p^* \tag{108}$$

$$\text{with } \pi_{p,n} \equiv \frac{\exp\Big(\beta_{0,p} + \sum_{j\in\{1,\ldots,K\}\setminus k} \beta_{j,p}x_j + \beta_{k,p}\bar{x}_{k,n}\Big)}{\sum_{o=1}^{P}\exp\Big(\beta_{0,o} + \sum_{j\in\{1,\ldots,K\}\setminus k} \beta_{j,o}x_j + \beta_{k,o}\bar{x}_{k,n}\Big)} \quad \forall\, p = 1,\ldots,P; n \in \{l, r\}, \tag{109}$$

$x_0 \equiv 1$, and $\Delta_{o,p}$ denoting Kronecker's Delta with $\Delta_{o,p} = 1 \;\forall\, o = p$ and $\Delta_{o,p} = 0 \;\forall\, o \neq p$.

If the regression equation additionally includes a quadratic term of the covariate of interest, we have:

$$\pi_{p,n} \equiv \frac{\exp\Big(\beta_{0,p} + \sum_{j\in\{1,\ldots,K\}\setminus\{k,k+1\}} \beta_{j,p}x_j + \beta_{k,p}\bar{x}_{k,n} + \beta_{k+1,p}\overline{x^2}_{k,n}\Big)}{\sum_{o=1}^{P}\exp(\beta_{0,o} + \sum_{j\in\{1,\ldots,K\}\setminus\{k,k+1\}} \beta_{j,o}x_j + \beta_{k,o}\bar{x}_{k,n} + \beta_{k+1,o}\overline{x^2}_{k,n})} \quad \forall\, p = 1,\ldots,P; n \in \{l, r\} \tag{110}$$

and there are $P\,(P-1)$ additional gradients:

$$\frac{\partial E_{k,lr,p}}{\partial \beta_{k+1,o}} = (\pi_{p,r}\,\pi_{o,r} - \Delta_{o,p}\,\pi_{p,r})\,\bar{x}^2_{kr} - (\pi_{p,l}\,\pi_{o,l} - \Delta_{o,p}\,\pi_{p,l})\,\bar{x}^2_{kl} \quad \forall\, p = 1,\ldots,P; o \in \{1,\ldots,P\}\setminus p^*. \tag{111}$$

A.5. Gradients of grouped and re-based effects of categorical and interval-coded covariates

A.5.1. Linear probability model

The gradients are:

$$\frac{\partial E_{k,lr}}{\partial \beta_j} = 0 \;\forall\; j \in \{0, \dots, K\} \setminus k \tag{112}$$

$$\frac{\partial E_{k,lr}}{\partial \delta_m} = D_{ml} - D_{mr} \;\forall\; m \in \{1, \dots, M\} \setminus m^* \tag{113}$$

A.5.2. Probit regression

The gradients are:

$$\frac{\partial E_{k,lr}}{\partial \beta_j} = \left(\phi_l(\cdot) - \phi_r(\cdot)\right) x_j \;\forall\; j \in \{0, \dots, K\} \setminus k \tag{114}$$

$$\frac{\partial E_{k,lr}}{\partial \delta_m} = \phi_l(\cdot)\; D_{ml} - \phi_r(\cdot)\; D_{mr} \;\forall\; m \in \{1, \dots, M\} \setminus m^* \tag{115}$$

$$\text{with } \phi_n(\cdot) \equiv \phi\left(\beta_0 + \sum_{j \in \{1,\dots,K\} \setminus k} \beta_j x_j + \sum_{m=1}^{M} \delta_m D_{mn}\right) \quad \forall\; n \in \{r, l\} \tag{116}$$

and $x_0 \equiv 1$.

A.5.3. Logistic regression

The gradients are:

$$\frac{\partial E_{k,lr}}{\partial \beta_j} = \left(\frac{\exp_l(\cdot)}{\left(1 + \exp_l(\cdot)\right)^2} - \frac{\exp_r(\cdot)}{\left(1 + \exp_r(\cdot)\right)^2}\right) x_j \quad \forall\; j \in \{0, \dots, K\} \setminus k \tag{117}$$

$$\frac{\partial E_{k,lr}}{\partial \delta_m} = \frac{\exp_l(\cdot)}{\left(1 + \exp_l(\cdot)\right)^2}\; D_{ml} - \frac{\exp_r(\cdot)}{\left(1 + \exp_r(\cdot)\right)^2}\; D_{mr} \quad \forall\; m \in \{1, \dots, M\} \setminus m^* \tag{118}$$

$$\text{with } \exp_n(\cdot) \equiv \exp\left(\beta_0 + \sum_{j \in \{1,\dots,K\} \setminus k} \beta_j x_j + \sum_{m=1}^{M} \delta_m D_{mn}\right) \quad \forall\; n \in \{r, l\} \tag{119}$$

and $x_0 \equiv 1$.

A.5.4. Multinomial logistic regression

The gradients are:

$$\frac{\partial E_{k,lr,\mathscr{P}}}{\partial \beta_{j,o}} = \sum_{p\in\mathscr{P}} \frac{\partial E^*_{k,lr,p}}{\partial \beta_{j,o}} \;\forall\, j \in \{0,\ldots,K\}\setminus k; o \in \{1,\ldots,P\}\setminus p^* \tag{120}$$

$$\frac{\partial E_{k,lr,\mathscr{P}}}{\partial \beta_{j,o}} = \sum_{p\in\mathscr{P}} \frac{\partial E^*_{k,lr,p}}{\partial \delta_{m,o}} \;\forall\, m \in \{1,\ldots,M\}\setminus m^*; o \in \{1,\ldots,P\}\setminus p^* \tag{121}$$

$$\text{with } \frac{\partial E^*_{k,lr,p}}{\partial \beta_{j,o}} = \left(\pi_{p,r}\;\pi_{o,r} - \pi_{p,l}\;\pi_{o,l} - \Delta_{o,p}\left(\pi_{p,r} - \pi_{p,l}\right)\right) x_j \tag{122}$$

$$\forall\, j \in \{0,\ldots,K\}\setminus k; p = 1,\ldots,P,$$

$$\frac{\partial E^*_{k,lr,p}}{\partial \delta_{m,o}} = \left(\pi_{p,r}\;\pi_{o,r} - \Delta_{o,p}\;\pi_{p,r}\right) D_{mr} - \left(\pi_{p,l}\;\pi_{o,l} - \Delta_{o,p}\;\pi_{p,l}\right) D_{ml} \tag{123}$$

$$\forall\, m \in \{1,\ldots,M\}\setminus m^*; p = 1,\ldots,P,$$

$$\pi_{p,n} \equiv \frac{\exp\left(\beta_{0,p} + \sum_{j\in\{1,\ldots,K\}\setminus k} \beta_{j,p} x_j + \sum_{m=1}^{M} \delta_{m,p} D_{m,n}\right)}{\sum_{o=1}^{P} \exp\left(\beta_{0,o} + \sum_{j\in\{1,\ldots,K\}\setminus k} \beta_{j,o} x_j + \sum_{m=1}^{M} \delta_{m,o} D_{m,n}\right)} \tag{124}$$

$$\forall\, p = 1,\ldots,P; n \in \{l,r\},$$

and $x_0 \equiv 1$.

B. Derivation of a binary probit model from an ordered probit model

$$
\begin{aligned}
\Pr(Y^* = 1|X = x) &= \Pr(Y \in \{p^*, \dots, P\}|X = x) && (125)\\
&= \sum_{p=p^*}^{P} \Pr(Y = p|X = x) && (126)\\
&= \sum_{p=p^*}^{P} \left(\Phi\left(\mu_p - \sum_{j=1}^{K} \beta_j x_j\right) - \Phi\left(\mu_{p-1} - \sum_{j=1}^{K} \beta_j x_j\right)\right) && (127)\\
&= \sum_{p=p^*}^{P} \Phi\left(\mu_p - \sum_{j=1}^{K} \beta_j x_j\right) - \sum_{p=p^*}^{P} \Phi\left(\mu_{P-1} - \sum_{j=1}^{K} \beta_j x_j\right) && (128)\\
&= \sum_{p=p^*}^{P} \Phi\left(\mu_p - \sum_{j=1}^{K} \beta_j x_j\right) - \sum_{p=p^*-1}^{P-1} \Phi\left(\mu_p - \sum_{j=1}^{K} \beta_j x_j\right) && (129)\\
&= \Phi\left(\mu_P - \sum_{j=1}^{K} \beta_j x_j\right) - \Phi\left(\mu_{p^*-1} - \sum_{j=1}^{K} \beta_j x_j\right) && (130)\\
&= \Phi\left(\infty\right) - \Phi\left(\mu_{p^*-1} - \sum_{j=1}^{K} \beta_j x_j\right) && (131)\\
&= 1 - \Phi\left(\mu_{p^*-1} - \sum_{j=1}^{K} \beta_j x_j\right) && (132)\\
&= \Phi\left(-\mu_{p^*-1} + \sum_{j=1}^{K} \beta_j x_j\right) && (133)
\end{aligned}
$$

C. Additional R code

C.1. Loading and preparing data

Loading the data set:

```
data( "Mroz87", package = "sampleSelection" )
```

Creating a dummy variable for the presence of children in the household:

```
Mroz87$kids <- Mroz87$kids5 + Mroz87$kids618
```

Creating dummy variables for interval-coding variable `age`:

```
Mroz87$age30.37 <- Mroz87$age >= 30 & Mroz87$age <= 37
Mroz87$age38.44 <- Mroz87$age >= 38 & Mroz87$age <= 44
Mroz87$age45.52 <- Mroz87$age >= 45 & Mroz87$age <= 52
Mroz87$age53.60 <- Mroz87$age >= 53 & Mroz87$age <= 60
all.equal(
  Mroz87$age30.37 + Mroz87$age38.44 + Mroz87$age45.52 + Mroz87$age53.60,
  rep( 1, nrow( Mroz87 ) ) )
```

Creating an ordered categorical variable that indicates three levels of labour force participation:

```
Mroz87$lfp3 <- factor( ifelse( Mroz87$hours == 0, "no",
  ifelse( Mroz87$hours <= 1300, "part", "full" ) ),
  ordered = TRUE, levels = c( "no", "part", "full" ) )
```

C.2. Probit regressions with `age` as linear covariate and with `age` as linear and quadratic covariate

Estimations and creating vectors with mean values of covariates:

```
estProbit <- glm( lfp ~ kids + age + educ,
  family = binomial(link = "probit"), data = Mroz87 )
xMean <- c( 1, colMeans( Mroz87[ , c( "kids", "age", "educ" ) ] ) )

estProbitQ <- glm( lfp ~ kids + age + I(age^2) + educ,
  family = binomial(link = "probit"), data = Mroz87 )
xMeanQ <- c( xMean[ 1:3], xMean[3]^2, xMean[4] )
```

C.3. Logistic regressions with age as interval-coded covariate

Estimation and creating a vector with mean values of covariates:

```
estLogitInt <- glm( lfp ~ kids + age30.37 + age38.44 + age53.60 + educ,
  family = binomial(link = "logit"), data = Mroz87 )
xMeanInt <- c( xMean[1:2], mean( Mroz87$age30.37 ),
  mean( Mroz87$age38.44 ), mean( Mroz87$age53.60 ), xMean[4] )
```

C.4. Ordered probit regression with age as linear and quadratic covariate

Estimation:

```
library( "MASS" )
estOProbitQ <- polr( lfp3 ~ kids + age + I(age^2) + educ,
  data = Mroz87, method = "probit", Hess = TRUE )
xMeanOProbit <- c( xMeanQ, -1 )
```

C.5. Multinomial logistic regressions with age as interval-coded covariate

Estimation:

```
library( "mlogit" )
estMLogitInt <- mlogit(
  lfp3 ~ 0 | kids + age30.37 + age38.44 + age53.60 + educ,
  data = Mroz87, reflevel = "no", shape = "wide" )
```