SFB
823

On interaction effects: The case of heckit and two-part models

Manuel Frondel, Colin Vance

Nr. 1/2012

Discussion Paper

SFB
823

On Interaction Effects: The Case of Heckit and Two-Part Models

Manuel Frondel, RWI and Ruhr University Bochum (RUB) and **Colin Vance**, RWI and Jacobs University Bremen

Abstract. Interaction effects capture the impact of one explanatory variable x_1 on the marginal effect of another explanatory variable x_2. To explore interaction effects, so-called interaction terms $x_1 x_2$ are typically included in estimation specifications. While in linear models the effect of a marginal change in the interaction term is equal to the interaction effect, this equality generally does not hold in non-linear specifications (AI, NORTON, 2003). This paper provides for a general derivation of interaction effects in both linear and non-linear models and calculates the formulae of the interaction effects resulting from HECKMAN's sample selection model as well as the Two-Part Model, two regression models commonly applied to data with a large fraction of either missing or zero values in the dependent variable, respectively. Drawing on a survey of automobile use from Germany, we argue that while it is important to test for the significance of interaction effects, their size conveys limited substantive content. More meaningful, and also more easy to grasp, are the conditional marginal effects pertaining to two variables that are assumed to interact.

JEL classification: C34.

Key words: Truncated regression models, interaction terms.

Correspondence: Prof. Dr. Manuel Frondel, Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI), Hohenzollernstr. 1-3, D-45128 Essen. E-mail: frondel@rwi-essen.de.

Acknowledgements: We are very grateful for invaluable comments and suggestions by Christoph M. SCHMIDT. This work has been supported by the Collaborative Research Center "Statistical Modelling of Nonlinear Dynamic Processes" (SFB 823) of the German Research Foundation (DFG), within the framework of Project A3, "Dynamic Technology Modelling".

1 Introduction

To explore whether the effect of an explanatory variable x_1 on the expected value $E[y]$ of the dependent variable y depends on the size of another explanatory variable x_2, it is indispensable to estimate the interaction effect, which is formally given by the second derivative $\frac{\partial^2 E[y]}{\partial x_2 \partial x_1}$. To this end, linear estimation specifications typically include so-called interaction terms, consisting of the product $z := x_1 x_2$ of two explanatory variables. In linear contexts, the marginal effect $\frac{\partial E[y]}{\partial(x_1 x_2)}$ of the interaction term $x_1 x_2$ equals the interaction effect $\frac{\partial^2 E[y]}{\partial x_2 \partial x_1}$. This equality, however, generally does not extend to non-linear specifications, as is demonstrated by AI and NORTON (2003) for the example of probit and logit models. Furthermore, NORTON, WANG, and AI (2004) emphasize that in non-linear models, interaction effects are generally conditional on all explanatory variables, rather than being constant, as in the linear case.

The present paper builds on the work of these authors in two respects. First, we calculate the formulae of the interaction effects resulting from HECKMAN's sample selection model, as well as the Two-Part Model, two commonly employed approaches to accommodate missing or zero values in the dependent variable, respectively.[1] Second, using an empirical example that applies both model types to travel survey data collected from a sample of motorists in Germany, we illustrate several subtleties inherent to the substantive interpretation of interaction effects gleaned from non-linear models. Most notably, we argue that while testing the statistical significance of an interaction effect is important, the economic content of its size is limited. In this regard, our discussion is perfectly in line with a recent article by GREENE (2010), who points out that, apart from statistical significance, one should care about economic and policy significance.

[1]Note that no canned programm is available for calculating the interaction effects resulting from the Two-Part Model, whereas the most recent version of Stata can calculate interaction terms for the HECKMAN model. Yet, Stata code does not handle cases when a variable appears in multiple interaction terms. To fill this void, our Stata Do-files that allow for replicating our empirical example are available upon request.

The following section provides for a general derivation of interaction effects for both linear and non-linear models. Section 3 presents a concise comparison of the Two-Part and HECKMAN model. Sections 4 and 5 derive the specific formulae of the interaction effects of the Heckit and the Two-Part model, followed by the presentation of an example in Section 6. The last section summarizes and concludes.

2 Interaction Effects

To provide a general derivation of interaction effects in both linear and non-linear models, we closely follow AI and NORTON (2003) and NORTON, WANG, and AI (2004).

2.1 Linear Models

We begin by drawing on the following linear specification of the expected value of dependent variable y:

$$E := E[y|x_1, x_2, \mathbf{w}] = \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \mathbf{w}^T \boldsymbol{\beta}, \tag{1}$$

where the parameters $\beta_1, \beta_2, \beta_{12}$, as well as the vector $\boldsymbol{\beta}$ are unknown and vector $\mathbf{w}$ excludes x_1 and x_2. Likewise, β_1, β_2, and β_{12} are excluded from vector $\boldsymbol{\beta}$.

Assuming that x_1 and x_2 are continuous variables, the marginal effect of x_1 on the expected value E is dependent on x_2 if $\beta_{12} \neq 0$:

$$\frac{\partial E}{\partial x_1} = \beta_1 + \beta_{12} x_2. \tag{2}$$

The impact of a marginal change in x_2 on the marginal effect of x_1, in other words the interaction effect, is then obtained from taking the derivative of (2) with respect to x_2:

$$\frac{\partial^2 E}{\partial x_2 \partial x_1} = \beta_{12}. \tag{3}$$

In linear specifications, therefore, the interaction effect $\frac{\partial^2 E}{\partial x_2 \partial x_1}$ equals the marginal effect $\frac{\partial E}{\partial (x_1 x_2)}$ of the interaction term $x_1 x_2$. For non-linear models, however, this equality generally does not hold, as is demonstrated in the subsequent section.

2.2 Non-Linear Models

Instead of expectation (1), we now depart from

$$E := E[y|x_1, x_2, \mathbf{w}] = F(\beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \mathbf{w}^T \boldsymbol{\beta}) = F(u), \tag{4}$$

where $F(u)$ is a non-linear function of its argument $u := \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \mathbf{w}^T \boldsymbol{\beta}$. In the Probit model, for example, $F(u)$ equals the cumulative normal distribution $\Phi(u)$. We now derive general formulae for the interaction effects resulting from non-linear models if (i) x_1 and x_2 are both continuous variables, (ii) both are dummy variables, and (iii) x_1 is continuous, while x_2 is a dummy variable.

(i) If $F(u)$ is a twice differentiable function, with the first and second derivatives being denoted by $F'(u)$ and $F''(u)$, respectively, the marginal effect with respect to x_1 reads:

$$\frac{\partial E}{\partial x_1} = \frac{\partial F(u)}{\partial x_1} = F'(u) \frac{\partial u}{\partial x_1} = F'(u)(\beta_1 + \beta_{12} x_2), \tag{5}$$

while the interaction effect of two continuous variables x_1 and x_2 is symmetric and given by

$$\frac{\partial^2 E}{\partial x_2 \partial x_1} = \frac{\partial}{\partial x_2}\left(\frac{\partial E}{\partial x_1}\right) = \frac{\partial}{\partial x_2}\left(\frac{\partial F(u)}{\partial x_1}\right) = F'(u)\beta_{12} + (\beta_1 + \beta_{12} x_2)(\beta_2 + \beta_{12} x_1) F''(u). \tag{6}$$

As, in general, $(\beta_1 + \beta_{12} x_2)(\beta_2 + \beta_{12} x_1) F''(u) \neq 0$, the interaction effect $\frac{\partial^2 E}{\partial x_2 \partial x_1}$ generally differs from the marginal effect $\frac{\partial E}{\partial (x_1 x_2)}$ of the interaction term $z = x_1 x_2$:

$$\frac{\partial E}{\partial (x_1 x_2)} = \frac{\partial E}{\partial z} = F'(u) \frac{\partial u}{\partial z} = F'(u) \beta_{12}. \tag{7}$$

(ii) If x_1 and x_2 are dummy variables, the discrete interaction effect, which in analogy to $\frac{\partial^2 E}{\partial x_2 \partial x_1}$ shall be designated by $\frac{\Delta^2 E}{\Delta x_2 \Delta x_1}$, is given by the discrete change in E due to a unitary change in both x_1 and x_2, $\Delta x_1 = 1, \Delta x_2 = 1$:

$$\begin{aligned} \frac{\Delta^2 E}{\Delta x_2 \Delta x_1} &:= \frac{\Delta}{\Delta x_2}\left(\frac{\Delta E}{\Delta x_1}\right) = \frac{\Delta}{\Delta x_2}\left(E[y|x_1 = 1, x_2, \mathbf{w}] - E[y|x_1 = 0, x_2, \mathbf{w}]\right) \\ &= \{E[y|x_1 = 1, x_2 = 1, \mathbf{w}] - E[y|x_1 = 0, x_2 = 1, \mathbf{w}]\} \\ &\quad -\{E[y|x_1 = 1, x_2 = 0, \mathbf{w}] - E[y|x_1 = 0, x_2 = 0, \mathbf{w}]\}. \end{aligned} \tag{8}$$

Note that the discrete interaction effects are symmetric: $\frac{\Delta^2 E}{\Delta x_2 \Delta x_1} = \frac{\Delta^2 E}{\Delta x_1 \Delta x_2}$, as can be seen from (8) by rearranging the terms in the middle of the double difference. Using the non-linear representation of expected value (4), the general expression (8) translates into:

$$\frac{\Delta^2 E}{\Delta x_2 \Delta x_1} = \{F(\beta_1 + \beta_2 + \beta_{12} + \mathbf{w}^T \boldsymbol{\beta})) - F(\beta_2 + \mathbf{w}^T \boldsymbol{\beta})\} - F(\beta_1 + \mathbf{w}^T \boldsymbol{\beta}) + F(\mathbf{w}^T \boldsymbol{\beta}). \quad (9)$$

(iii) If x_1 is a continuous variable and x_2 is a dummy variable, the mixed interaction effect $\frac{\Delta}{\Delta x_2}(\frac{\partial E}{\partial x_1})$ can be computed on the basis of the marginal effect (5) as follows:

$$\begin{aligned} \frac{\Delta}{\Delta x_2}\left(\frac{\partial E}{\partial x_1}\right) &:= \frac{\Delta}{\Delta x_2}\left(\frac{\partial F(u)}{\partial x_1}\right) = \left.\frac{\partial F(u)}{\partial x_1}\right|_{x_2=1} - \left.\frac{\partial F(u)}{\partial x_1}\right|_{x_2=1} \qquad (10) \\ &= F'(\beta_1 x_1 + \beta_2 + \beta_{12} x_1 + \mathbf{w}^T \boldsymbol{\beta})(\beta_1 + \beta_{12}) - F'(\beta_1 x_1 + \mathbf{w}^T \boldsymbol{\beta})\beta_1. \end{aligned}$$

The symmetry observed for the cases when both variables are either continuous or dummies also holds true for the mixed interaction effects: $\frac{\partial}{\partial x_1}(\frac{\Delta E}{\Delta x_2}) = \frac{\Delta}{\Delta x_2}(\frac{\partial E}{\partial x_1})$.[2]

All in all, it bears noting that for linear functions such as $F(u) = u$, for which $F'(u) = 1$, all three kinds of interaction effects collapse to β_{12}. Furthermore, we shall re-emphasize the point raised by AI and NORTON (2003:124) that, in contrast to linear specifications, the interaction effect gleaned from non-linear models is generally non-vanishing even if no interaction term is included, that is, if $\beta_{12} = 0$.

Finally, for the special case of the Probit model, the interaction effects are given by (6), (9), and (10) if $F(u)$ is replaced by the cumulative standard normal distribution $\Phi(u)$, $F'(u)$ is replaced by the density function of the standard normal distribution, $\phi(u) := \exp\{-u^2/2\}/\sqrt{2\pi}$, and $F''(u)$ is replaced by $\phi'(u) = -u\phi(u)$. Similarly, formulae (6), (9), and (10) can be applied to the Logit model if $F(u)$ is replaced by $\Lambda(u) := 1/(1 + \exp\{-u\})$, $F'(u)$ is replaced by $\Lambda'(u) = \Lambda(u)(1 - \Lambda(u))$, and $F''(u)$ is substituted by $\Lambda''(u) = (\Lambda(u)(1 - \Lambda(u)))' = \Lambda(u)(1 - \Lambda(u))(1 - 2\Lambda(u))$.

[2] Yet, note that $\frac{\Delta}{\Delta x_2}(\frac{\partial E}{\partial x_1}) \neq \frac{\partial}{\partial x_2}(\frac{\Delta E}{\Delta x_1})$.

3 Two-Part and Heckit Models

To accommodate the feature of a large proportion of missing or zero values in dependent variables, two-stage estimation procedures such as HECKMAN's (1979) sample selection or the Two-Part model (2PM) are frequently employed. Confusion reigns, however, about the proper use of these models. Above all, this confusion originates from frequently incorrect interpretations of the zero values of the outcome variable. These zeros may indicate (1) exogenous censoring[3], where the dependent variable can, in principle, take on negative values, but instead only zeros and positive values are observed, (2) true zeros, as is the case for, say, automobile and health expenditures, or (3) missing data, as in the analysis of hours worked or wages in labor economics.

For a clear distinction, it is helpful to employ the nomenclature of *actual* versus *potential* outcomes introduced by DOW and NORTON (2003). In this terminology, the actual outcome designates a fully observed variable, with zero values for, say, health expenditures representing true zeros and indicating that actually no money for health care is expended. If instead zero values indicate observations for which values of the dependent variable, such as wages, are missing, these authors use the term potential outcome for this variable. It is a latent variable that is merely partially observed, in this instance when wages are positive. In the context of our empirical example presented in Section 6, the potential outcome addresses the distance an individual would potentially drive with a car irrespective of actual car use, while the actual outcome is the observed distance driven over a certain time period. It becomes immediately clear that in our example the actual outcome provides for the more natural interpretation than the potential outcome.

Although HECKMAN's sample selection model, which is frequently called Heckit, can be used to estimate actual outcomes, this interpretation requires several extra calculations beyond what is commonly provided by statistical software packages (DOW, NORTON, 2003:6). As an alternative, DUAN et al. (1984) proposed the 2PM, arguing that

[3] In this case, which is outside the scope of this article, the so-called Tobit model is the standard estimation procedure.

it often has lower mean squared error than the Heckit estimator when analyzing actual outcomes. In a similar vein, it is argued by DOW and NORTON (2003:6) that for this purpose, the Heckit incorporates features that make it often perform worse than the 2PM estimator, because the Heckit was designed to address selection bias in the analysis of potential rather than actual outcomes. On the other hand, the 2PM would suffer from selection bias if observations with zeros in the dependent variable to indicate missing values differ systematically from those with observed values (DOW, NORTON, 2003:6).

Before applying the Heckit and the 2PM two-stage estimation procedures, we shall briefly summarize the corresponding model structures. The Heckit orders observations of the dependent variable y into two regimes, where the first stage defines a dichotomous variable R, indicating the regime into which the observation falls:

$$R = 1, \text{if} \quad R^* = \mathbf{x_1}^T \boldsymbol{\tau} + \epsilon_1 > 0 \quad \text{and} \quad R = 0, \text{if} \quad R^* \leq 0, \tag{11}$$

where R^* is a latent variable, vector $\mathbf{x_1}$ includes its determinants, $\boldsymbol{\tau}$ is a vector of associated parameters, and ϵ_1 is an error term assumed to have a standard normal distribution. $R = 1$ indicates that $y > 0$, whereas otherwise $R = 0$ if values of y are either missing or missing values are denoted by zero. For the 2PM, the actual outcomes y can also be ordered into two regimes, with $R = 1$ denoting positive outcomes ($y > 0$), whereas $R = 0$ is equivalent to $y = 0$.

Both model types include a Probit estimation of the probability of having positive outcomes (regime $R = 1$) as the first stage of the two-stage estimation procedure (selection equation):

$$P(y > 0|\mathbf{x_1}) = \Phi(\mathbf{x_1}^T \boldsymbol{\tau}). \tag{12}$$

The second stage of both model types involves estimating the parameters $\boldsymbol{\beta}$ of interest via an OLS regression conditional on $R = 1$, i. e. $y > 0$ (conditional equation):

$$E[y|R = 1, \mathbf{x_2}] = E[y|y > 0, \mathbf{x_2}] = \mathbf{x_2}^T \boldsymbol{\beta} + E(\epsilon_2|y > 0, \mathbf{x_2}), \tag{13}$$

where $\mathbf{x_2}$ includes the determinants of the dependent variable y, and ϵ_2 is another error term.

The prediction of the dependent variable y consists of two parts, with the first part resulting from the first stage (12), $P(y > 0) = \Phi(\mathbf{x_1}^T\boldsymbol{\tau})$, and the second part being the conditional expectation $E[y|y > 0]$ from the second stage (13):

$$E[y] = P(y > 0) \cdot E[y|y > 0] + P(y = 0) \cdot E[y|y = 0] = P(y > 0) \cdot E[y|y > 0].$$

In the 2PM, where it is assumed that $E(\epsilon_2|y > 0, \mathbf{x_2}) = \mathbf{0}$ and, hence, $E[y|y > 0, \mathbf{x_2}] = \mathbf{x_2}^T\boldsymbol{\beta}$, the unconditional expectation $E[y]$ is given by:

$$E[y] = \Phi(\mathbf{x_1}^T\boldsymbol{\tau}) \cdot \mathbf{x_2}^T\boldsymbol{\beta}. \tag{14}$$

By contrast, the second stage OLS regression of the Heckit model includes the inverse MILLS ratio, $\lambda(\mathbf{x_1}^T\boldsymbol{\tau})) := \frac{\varphi(\mathbf{x_1}^T\boldsymbol{\tau})}{\Phi(\mathbf{x_1}^T\boldsymbol{\tau})}$, as an additional regressor to control for sample selectivity:[4]

$$E[y|y > 0] = \mathbf{x_2}^T\boldsymbol{\beta} + \beta_\lambda \cdot \lambda(\mathbf{x_1}^T\boldsymbol{\tau}), \tag{15}$$

where β_λ is called the sample-selection parameter and the inverse MILLS ratio is proportional to $E(\epsilon_2|y > 0, \mathbf{x_2}) \neq 0$ when ϵ_2 is assumed to be normally distributed with constant variance: $\mathrm{Var}(\epsilon_2) = \sigma^2$.

Finally, it bears noting that it is advisable to include so-called exclusion restrictions when estimating the Heckit model, implying that the sets of regressors $\mathbf{x_1}$ and $\mathbf{x_2}$ of both stages differ at least in one variable. This ensures that the model is well-identified, thereby avoiding multi-collinearity problems due to the inclusion of the inverse MILLS ratio in equation (15). In contrast, such exclusion restrictions are unnecessary in the 2PM, so that in practice both sets of regressors can be identical: $\mathbf{x_1} = \mathbf{x_2}$.

4 Interaction Effects in Heckit Models

The second stage of the Heckit model relies upon the **conditional** expectation

$$E = E[y|x_1, x_2, \mathbf{w_1}, \mathbf{w_2}, y > 0] = u_2 + \beta_\lambda \cdot \lambda(u_1), \tag{16}$$

[4]While the Heckit model consists of the two-stage estimation procedure described above, modern computer software has made the full information maximum likelihood (FIML) variant, referred to as the Heckman model, the most often used.

where $u_1 := \tau_1 x_1 + \tau_2 x_2 + \tau_{12} x_1 x_2 + \mathbf{w_1}^T \boldsymbol{\tau}$, $u_2 := \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \mathbf{w_2}^T \boldsymbol{\beta}$, and $\lambda(u_1) := \frac{\varphi(u_1)}{\Phi(u_1)}$ denotes the inverse MILLS ratio, β_λ is the respective coefficient, and $\mathbf{w_1}$ and $\mathbf{w_2}$ both exclude x_1 and x_2. Likewise, the parameters τ_1, τ_2 and τ_{12} are not included in vector $\boldsymbol{\tau}$, nor are β_1, β_2, and β_{12} part of vector $\boldsymbol{\beta}$. Finally, note that in both u_1 and u_2 we include the same interaction term $x_1 x_2$, while, of course, the interaction terms occurring in u_1 and u_2 could also be different.

Before deriving the formulae for the interaction effects, it should be recognized that

$$\lambda'(u_1) = \frac{-u_1 \varphi(u_1)\Phi(u_1) - \varphi^2(u_1)}{\Phi^2(u_1)} = -[\lambda(u_1)]^2 - u_1 \cdot \lambda(u_1)$$

and

$$\lambda''(u_1) = -2\lambda(u_1) \cdot \lambda'(u_1) - \lambda(u_1) - u_1 \cdot \lambda'(u_1) = -[2\lambda(u_1) + u_1] \cdot \lambda'(u_1) - \lambda(u_1).$$

(i) To calculate the interaction effect of two continuous variables, we first need to calculate the marginal effect:

$$\frac{\partial E}{\partial x_1} = (\beta_1 + \beta_{12} x_2) + \beta_\lambda \cdot \lambda'(u_1) \cdot (\tau_1 + \tau_{12} x_2). \tag{17}$$

Apparently, marginal effects resulting from non-linear models generally depend on all other variables. As elaborated in the empirical example below, the correct calculation of the marginal effect $\frac{\partial E}{\partial x_1}$ necessitates that the derivatives $\tau_{12} x_2$ and $\beta_{12} x_2$ of the interaction terms must be taken into account.

Note also that, as is pointed out by AI and NORTON (2003:123), it would be incorrect to calculate the interaction effect by taking the marginal effect of the interaction term $z = x_1 x_2$:

$$\frac{\partial E}{\partial z} = \beta_{12} + \beta_\lambda \cdot \lambda'(u_1) \cdot \tau_{12}. \tag{18}$$

The correct interaction effect can instead by obtained by taking the derivative of (17) with respect to x_2:

$$\frac{\partial^2 E}{\partial x_2 \partial x_1} = \beta_{12} + \beta_\lambda \cdot \{\lambda''(u_1) \cdot (\tau_2 + \tau_{12} x_1) \cdot (\tau_1 + \tau_{12} x_2) + \lambda'(u_1) \cdot \tau_{12}\}. \tag{19}$$

Note that the derivatives given by (18) and (19) are generally different and would only be identical if $\lambda''(u_1) \cdot (\tau_2 + \tau_{12}x_1) \cdot (\tau_1 + \tau_{12}x_2) = 0$.

(ii) On the basis of the marginal effect (17), the mixed interaction effect $\frac{\Delta}{\Delta x_2}(\frac{\partial E}{\partial x_1})$ is given by:

$$\begin{aligned}
\frac{\Delta}{\Delta x_2}\left(\frac{\partial E}{\partial x_1}\right) &= \left.\frac{\partial E}{\partial x_1}\right|_{x_2=1} - \left.\frac{\partial E}{\partial x_1}\right|_{x_2=0} \qquad (20)\\
&= (\beta_1 + \beta_{12}) + \beta_\lambda \cdot \lambda'(\tau_1 x_1 + \tau_2 + \tau_{12}x_1 + \mathbf{w_1}^T\boldsymbol{\tau}) \cdot (\tau_1 + \tau_{12})\\
&\quad -\beta_1 - \beta_\lambda \cdot \lambda'(\tau_1 x_1 + \mathbf{w_1}^T\boldsymbol{\tau}) \cdot \tau_1\\
&= \beta_{12} + \beta_\lambda \cdot \{\lambda'(\tau_1 x_1 + \tau_2 + \tau_{12}x_1 + \mathbf{w_1}^T\boldsymbol{\tau}) \cdot (\tau_1 + \tau_{12})\\
&\quad -\lambda'(\tau_1 x_1 + \mathbf{w_1}^T\boldsymbol{\tau}) \cdot \tau_1\}.
\end{aligned}$$

(iii) Using expectation (16), the discrete interaction effect reads as follows:

$$\begin{aligned}
\frac{\Delta^2 E}{\Delta x_2 \Delta x_1} &= \{[E[y|x_1 = 1, x_2 = 1, \mathbf{w_1}, \mathbf{w_2}] - E[y|x_1 = 0, x_2 = 1, \mathbf{w_1}, \mathbf{w_2}]\}\\
&\quad -\{[E[y|x_1 = 1, x_2 = 0, \mathbf{w_1}, \mathbf{w_2}] - E[y|x_1 = 0, x_2 = 0, \mathbf{w_1}, \mathbf{w_2}]\}\\
&= \beta_{12} + \beta_\lambda\{\lambda(\tau_1 + \tau_2 + \tau_{12} + \mathbf{w_1}^T\tau) - \lambda(\tau_2 + \mathbf{w_1}^T\tau) \qquad (21)\\
&\quad -\lambda(\tau_1 + \mathbf{w_1}^T\tau) + \lambda(\mathbf{w_1}^T\tau)\}.
\end{aligned}$$

Note that in all three cases the interaction effect collapses to the coefficient β_{12} of the interaction term if $\beta_\lambda = 0$, that is, when the inverse MILLS ratio is neglected and the Heckit model degenerates to the classical linear regression model.

5 Interaction Effects in Two-Part Models

In this section, we derive the formulae for the interaction effects resulting from Two-Part Models (2PM) for the special case that variable x_1 interacts with two, rather than only one other variable, as in the previous section. To this end, we use a more detailed version of the **unconditional** expectation (14),

$$E := E[y|x_1, x_2, x_3, \mathbf{w_1}, \mathbf{w_2}] = \Phi(u_1)u_2,$$

where now $u_1 := \tau_1 x_1 + \tau_2 x_2 + \tau_3 x_3 + \tau_{12} x_1 x_2 + \tau_{13} x_1 x_3 + \mathbf{w_1}^T \boldsymbol{\tau}$, $u_2 := \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \mathbf{w_2}^T \boldsymbol{\beta}$, and $\mathbf{w_1}$ and $\mathbf{w_2}$ neither include x_1, nor x_2 and x_3. Likewise, the parameters $\tau_1, \tau_2, \tau_3, \tau_{12}$, and τ_{13} are not included in vector $\boldsymbol{\tau}$, nor are $\beta_1, \beta_2, \beta_3, \beta_{12}$, and β_{13} part of vector $\boldsymbol{\beta}$. We derive formulae for the interaction effects if (i) x_1 and x_2 are both continuous variables, (ii) x_1 is continuous, while x_2 is a dummy variable, and (iii) both are dummy variables.

(i) To calculate the interaction effect $\frac{\partial^2 E}{\partial x_2 \partial x_1}$, we once again need to calculate the marginal effect:

$$\frac{\partial E}{\partial x_1} = (\tau_1 + \tau_{12} x_2 + \tau_{13} x_3) \cdot \varphi(u_1) \cdot u_2 + \Phi(u_1) \cdot (\beta_1 + \beta_{12} x_2 + \beta_{13} x_3). \tag{22}$$

By taking the derivative with respect to x_2 and employing $\varphi'(u_1) = -u_1 \varphi(u_1)$, we get the interaction effect of two continuous variables x_1 and x_2:

$$\begin{aligned} \frac{\partial^2 E}{\partial x_2 \partial x_1} &= \tau_{12} \cdot \varphi(u_1) \cdot u_2 - (\tau_1 + \tau_{12} x_2 + \tau_{13} x_3) \cdot (\tau_2 + \tau_{12} x_1) \cdot \varphi(u_1) \cdot u_1 \cdot u_2 \\ &\quad + (\tau_1 + \tau_{12} x_2 + \tau_{13} x_3) \cdot \varphi(u_1) \cdot (\beta_2 + \beta_{12} x_1) \\ &\quad + (\tau_2 + \tau_{12} x_1) \cdot \varphi(u_1) \cdot (\beta_1 + \beta_{12} x_2 + \beta_{13} x_3) + \Phi(u_1) \cdot \beta_{12}. \end{aligned} \tag{23}$$

(ii) The mixed interaction effect $\frac{\Delta}{\Delta x_2}(\frac{\partial E}{\partial x_1})$ follows immediately from the marginal effect (22):

$$\begin{aligned} \frac{\Delta}{\Delta x_2}\left(\frac{\partial E}{\partial x_1}\right) &= \left.\frac{\partial E}{\partial x_1}\right|_{x_2=1} - \left.\frac{\partial E}{\partial x_1}\right|_{x_2=0} \\ &= (\tau_1 + \tau_{12} + \tau_{13} x_3) \cdot \varphi(\tau_1 x_1 + \tau_2 + \tau_3 x_3 + \tau_{12} x_1 + \tau_{13} x_1 x_3 + \mathbf{w_1}^T \boldsymbol{\tau}) \cdot \\ &\quad \cdot \{\beta_1 x_1 + \beta_2 + \beta_3 x_3 + \beta_{12} x_1 + \beta_{13} x_1 x_3 + \mathbf{w_2}^T \boldsymbol{\beta}\} \\ &\quad + \Phi(\tau_1 x_1 + \tau_2 + \tau_3 x_3 + \tau_{12} x_1 + \tau_{13} x_1 x_3 + \mathbf{w_1}^T \boldsymbol{\tau}) \cdot (\beta_1 + \beta_{12} + \beta_{13} x_3) \\ &\quad - (\tau_1 + \tau_{13} x_3) \cdot \varphi(\tau_1 x_1 + \tau_3 x_3 + \tau_{13} x_1 x_3 + \mathbf{w_1}^T \boldsymbol{\tau}) \cdot \\ &\quad \cdot \{\beta_1 x_1 + \beta_3 x_3 + \beta_{13} x_1 x_3 + \mathbf{w_2}^T \boldsymbol{\beta}\} \\ &\quad - \Phi(\tau_1 x_1 + \tau_3 x_3 + \tau_{13} x_1 x_3 + \mathbf{w_1}^T \boldsymbol{\tau}) \cdot (\beta_1 + \beta_{13} x_3). \end{aligned} \tag{24}$$

(iii) Applying formula (8) to $E[y|x_1, x_2, x_3, \mathbf{w_1}, \mathbf{w_2}]$, the discrete interaction effect $\frac{\Delta^2 E}{\Delta x_2 \Delta x_1}$

is obtained as follows:

$$\begin{aligned}
\frac{\Delta^2 E}{\Delta x_2 \Delta x_1} &= \{[E[y|x_1 = 1, x_2 = 1, \mathbf{w_1}, \mathbf{w_2}] - E[y|x_1 = 0, x_2 = 1, \mathbf{w_1}, \mathbf{w_2}]\} \\
&\quad -\{[E[y|x_1 = 1, x_2 = 0, \mathbf{w_1}, \mathbf{w_2}] - E[y|x_1 = 0, x_2 = 0, \mathbf{w_1}, \mathbf{w_2}]\} \\
&= \Phi(\tau_1 + \tau_2 + \tau_3 x_3 + \tau_{12} + \tau_{13} x_3 + \mathbf{w_1}^T \boldsymbol{\tau}) \\
&\quad \cdot \{\beta_1 + \beta_2 + \beta_3 x_3 + \beta_{12} + \beta_{13} x_3 + \mathbf{w_2}^T \boldsymbol{\beta}\} \\
&\quad -\Phi(\tau_2 + \tau_3 x_3 + \mathbf{w_1}^T \boldsymbol{\tau}) \cdot \{\beta_2 + \beta_3 x_3 + \mathbf{w_2}^T \boldsymbol{\beta}\} \\
&\quad -\Phi(\tau_1 + (\tau_3 + \tau_{13}) x_3 + \mathbf{w_1}^T \boldsymbol{\tau}) \cdot \{\beta_1 + (\beta_3 + \beta_{13}) x_3 + \mathbf{w_2}^T \boldsymbol{\beta}\} \\
&\quad +\Phi(\tau_3 x_3 + \mathbf{w_1}^T \tau) \cdot \{\beta_3 x_3 + \mathbf{w_2}^T \boldsymbol{\beta}\}.
\end{aligned} \tag{25}$$

6 Empirical Example

To illustrate the estimation of the interaction effects gleaned from both a Heckit and a Two-Part Model (2PM), we employ household data drawn from the German Mobility Panel (MOP 2011) using the following specifications for the 2PM:

$$E[s] = \Phi(\mathbf{x_1}^T \boldsymbol{\tau}) \cdot \{\mathbf{x_2}^T \boldsymbol{\beta}\} \tag{26}$$

and for the Heckit model:

$$E[s] = \mathbf{x_2}^T \boldsymbol{\beta} + \beta_\lambda \cdot \frac{\phi(\mathbf{x_1}^T \boldsymbol{\tau})}{\Phi(\mathbf{x_1}^T \boldsymbol{\tau})}, \tag{27}$$

where the dependent variable s is the daily distance driven for non-work travel and the sets of explanatory variables $\mathbf{x_1}$ and $\mathbf{x_2}$ include the individual and household attributes that are hypothesized to influence the extent of this travel.

Note again that in the 2PM, the variable sets $\mathbf{x_1}$ and $\mathbf{x_2}$ can be identical, while for the Heckit model they must differ in at least one variable (exclusion restriction), as otherwise the Heckit would solely be identified through its non-linearity. In our empirical example, we include three variables that serve to satisfy the exclusion restriction: a dummy indicating whether the household has access to a *private parking* space, a continuous measure of walking *minutes* to the nearest public transit stop, and a dummy indicating whether this stop is serviced by *rail transit* (as opposed to bus). Because each of these variables is intended to capture the fixed costs of forgoing automobile travel, they are included in the (1. stage) Probit model of automobile use, but excluded from the (2. stage) estimation of distance traveled. Variable definitions and descriptive statistics are presented in Table A1 in Appendix A. A detailed data description can be found in FRONDEL, PETERS, VANCE (2008) or FRONDEL, VANCE (2009, 2010).

The key attributes of interest in our example are the individual's *age*, the number (#) of *children*, and the dummy variable *enoughcars* indicating whether the individual lives in a household in which the number of cars is at least equal to the number of licensed drivers. These variables are of particular interest because of major demographic and socioeconomic changes underway in Germany whose implications for transportation are potentially profound. By 2050, for example, Germany's population is projected to shrink by roughly 16% (STABUA, 2006), a trend that will be paralleled by an increasingly older age structure of the German population. At the same time, the prevalence of automobiles in Germany has been steadily rising; between 2002 and 2007 the number of privately owned automobiles increased 5%, from 39.6 to 41.6 million (KBA, 2011). To explore how the role of demographics and car availability are mediated by gender in dictating access to and use of the car, these variables are interacted with a *female* dummy variable.

Table 1 reports the results from a Heckit model for two model specifications, one in which several interaction terms are included and another in which these are omitted entirely. To focus on the salient results, we refrain here from reporting the estimation results of the (1. stage) Probit model and instead present both the coefficient estimates of the (2. stage) OLS regression, as well as the marginal and interaction effects of the explanatory variables on distance driven resulting from the Heckit model. The presented estimates of the Heckit are based on the calculation of the mean of the individual marginal effects for each observation in the data. Moreover, given that the marginal and interaction effects are comprised of multiple parameters that make analytical computation of the variance impossible, bootstrapping was used to calculate the standard errors.[5]

Table 1: Heckit Estimation Results on Distance driven.

	Interaction Terms Included: $\tau_{12} \neq 0, \beta_{12} \neq 0$				No Interaction Terms: $\tau_{12} = \beta_{12} = 0$			
	Coeff.s	Errors	Effects	Errors	Coeff.s	Errors	Effects	Errors
female	*-1.280	(0.577)	**-1.614	(0.201)	**-1.571	(0.302)	** -1.515	(0.208)
employed	-0.222	(0.618)	**-1.285	(0.355)	-0.221	(0.592)	** -1.168	(0.346)
commute distance	** 0.040	(0.008)	** 0.289	(0.006)	** 0.041	(0.008)	** 0.029	(0.006)
age	**-0.066	(0.020)	**-0.037	(0.011)	**-0.061	(0.018)	** -0.038	(0.011)
age $\times$ *# children*	0.013	(0.019)	0.005	(0.016)	–	–	** -0.0012	(0.0005)
high-school diploma	** 1.000	(0.318)	** 1.055	(0.270)	** 1.003	(0.315)	** 1.088	(0.280)
# children	*-1.833	(0.981)	-0.218	(0.157)	**-1.277	(0.457)	* -0.262	(0.132)
# employed	-0.099	(0.245)	-0.048	(0.199)	-0.091	(0.242)	-0.077	(0.197)
enoughcars	-0.178	(0.570)	** 1.390	(0.233)	-0.397	(0.776)	** 1.365	(0.235)
female $\times$ *enoughcars*	-0.569	(0.901)	** 1.276	(0.436)	–	–	-0.005	(0.006)
city region	**-0.942	(0.301)	**-0.661	(0.222)	**-0.933	(0.302)	** -0.660	(0.221)
Inverse MILLS *ratio*	**-7.840	(3.214)	–	–	**-7.569	(3.196)	–	–
# observations used for estimation: 44,842								

Note: * denotes significance at the 5 %-level and ** at the 1 %-level, respectively. Estimates of the marginal and interaction effects are average effects that have been computed by averaging the marginal effects over the observations.

Turning first to the model that includes the interaction terms, the OLS estimates and associated marginal effects of the Heckit are seen to differ markedly, both with respect to their magnitude and statistical significance. For some of the variables, such as *employed* and *enoughcars*, statistically significant estimates of the marginal effects correspond to insignificant coefficient estimates, while for # *children* the converse is true. Apparently, testing the hypothesis that a marginal effect equals zero is not equivalent to the hypothesis that the variable in question is not a statistically significant determinant of the outcome. This is due to the fact that marginal effects are a non-linear function of all the coefficients of the model. Given this distinction, it bears noting that

[5]An alternative method for calculating the standard error, which employs a Taylor expansion, is the Delta method; see Vance (2009) for a comparison of the Delta method with bootstrapping.

GREENE (2010: 292, 2007: E18-23) argues that significance tests should be based on the coefficients, rather than on the marginal effects.

In a similar vein, it appears to be particularly important to distinguish between interaction terms and interaction effects: For example, while the estimate of the coefficient of the interaction term *female* $\times$ *enoughcars* does not statistically differ from zero, the associated interaction effect is significantly positive. Although no interaction terms are included in the specification presented on the right-hand panel, at least in one case, for the interaction of *age* and # *children*, the corresponding interaction effect, which is calculated using formula (19) and setting $\tau_{12} = \beta_{12} = 0$, is still significantly different from zero. This serves to highlight the fact that the marginal effect of a variable x_1 depends on variable x_2, even when no interaction term $x_1 x_2$ is included in the model.

Moreover, we now illustrate that the statistical significance of interaction effects warrants testing. For example, with *enoughcars*=1 designating that there are at least as many cars as licensed drivers in a household, the interaction effect of 1.276 of the dummy variables *female* and *enoughcars* indicates a statistically significant difference of the conditional marginal effects of a sufficient versus an insufficient number of cars among male and female persons, and hence signals gender competition for cars. More generally, the interaction effect shows how the partial effect of a variable x_1, such as the binary variable *enoughcars*, varies with a change in another variable x_2, for instance a regime switch in the gender variable *female*, and vice versa. Despite this straightforward qualitative interpretation, though, the size of an interaction effect is hard to grasp.

A key reason is that the interaction effect may be split up in either of two ways with equal justification. The first way involves calculating the impact of sufficient cars among females and males. For females, this is given by:

$$\begin{aligned}\frac{\Delta E}{\Delta enoughcars}|_{female=1} \quad &= \quad E[y|enoughcars = 1, female = 1, \mathbf{w_1}, \mathbf{w_2}] \\ &\quad -E[y|enoughcars = 0, female = 1, \mathbf{w_1}, \mathbf{w_2}] = 2.059^{**},\end{aligned}$$

and for males by:

$$\frac{\Delta E}{\Delta enoughcars}|_{female=0} = E[y|enoughcars = 1, female = 0, \mathbf{w_1}, \mathbf{w_2}]$$
$$-E[y|enoughcars = 0, female = 0, \mathbf{w_1}, \mathbf{w_2}] = 0.783,$$

where asterisks indicate that the first conditional marginal effect is statistically significant at the 1% level. The difference of this pair of conditional marginal effects, which equals the interaction effect of 1.276 reported for the variables *female* and *enoughcars* on the left-hand panel of Table 1, differs from zero, as this interaction effect $\frac{\Delta^2 E}{\Delta enough \Delta female}$ is non-vanishing and statistically different from zero according to Table 1.

On the other hand, given that in this instance we are dealing with a double difference, the same interaction effect of 1.276 also results from the difference of the following two marginal effects: first, the statistically significant marginal effect

$$\frac{\Delta E}{\Delta female}|_{enoughcars=1} = [E[y|female = 1, enoughcars = 1, \mathbf{w_1}, \mathbf{w_2}]$$
$$-E[y|female = 0, enoughcars = 1, \mathbf{w_1}, \mathbf{w_2}] = -1.065^{**},$$

which indicates that among households with a sufficient number of cars, there are significant differences between female and male car use for non-work purposes. Moreover, in households with less cars than licensed drivers, females drive 2.34 non-work kilometers less per day than males, confirming a large body of literature on gender differences in mobility behavior (e.g. WHITE, 1986; LEE, MCDONALD, 2003; MCDONALD, 2005):

$$\frac{\Delta E}{\Delta female}|_{enoughcars=0} = [E[y|female = 1, enoughcars = 0, \mathbf{w_1}, \mathbf{w_2}]$$
$$-E[y|female = 0, enoughcars = 0, \mathbf{w_1}, \mathbf{w_2}] = -2.340^{**}.$$

In short, we have exemplified that useful quantitative interpretations can be gleaned from breaking the interaction effect into its constituent parts and testing the statistical significance of each conditional marginal effect.

When comparing the empirical results obtained from the Heckit and the Two Part Model (2PM), it cannot be emphasized enough that both models address distinct research questions. The marginal effects of the Heckit, for instance, which are derived from conditional expectation (27) that incorporates the inverse MILLS ratio, are to be interpreted as the explanatory variables' impact on potential outcomes (see DOW and NORTON, 2003). Hence, because their interpretation is fundamentally different, we should not expect similar results for the Heckit and the 2PM. Indeed, in several cases, e. g. for the variables *commute distance* and #*children*, the Heckit outcomes lead to qualitatively different conclusions than those pertaining to the effects on actual outcomes from 2PM, which are reported in Table 2.[6]

In sum, our empirical example demonstrates how both models, the Heckit and the 2PM, are able to capture the conceptually subtle issue emerging from the fact that, in daily travel behavior, some motorists choose not to use their car, and whose recorded driving therefore equals zero. If ignored, the presence of these null values in the data is shown to potentially result in spurious conclusions with respect to both the magnitude and the significance of the estimates on car mileage (FRONDEL and VANCE, 2009).

7 Summary and Conclusion

By providing a general derivation of interaction effects in both linear and non-linear models and the specific formulae of the interaction effects gleaned from the Heckit and the Two-Part Model (2PM), this paper has analyzed the significance of these effects. Drawing on a survey of automobile use from Germany, we have illustrated that a non-

[6]Deviating from the formulae provided in Section 5, the interaction effect of the *female* dummy and the variable for the number of children is calculated as follows, where x_1 = *female* and x_3 = # *children*:

$$
\begin{aligned}
\frac{\Delta}{\Delta x_1}\left(\frac{\partial E}{\partial x_3}\right) &= \left.\frac{\partial E}{\partial x_3}\right|_{x_1=1} - \left.\frac{\partial E}{\partial x_3}\right|_{x_1=0} \\
&= (\tau_3+\tau_{13})\cdot\varphi(\tau_1+\tau_2x_2+\tau_3x_3+\tau_{12}x_2+\tau_{13}x_3+\mathbf{w_1}^T\boldsymbol{\tau})\cdot \\
&\quad \cdot\{\beta_1+\beta_2x_2+\beta_3x_3+\beta_{12}x_2+\beta_{13}x_3+\mathbf{w_2}^T\boldsymbol{\beta}\} \\
&\quad +\Phi(\tau_1+\tau_2x_2+\tau_3x_3+\tau_{12}x_2+\tau_{13}x_3+\mathbf{w_1}^T\boldsymbol{\tau})\cdot(\beta_1+\beta_{13}) \qquad (28) \\
&\quad -\tau_3\cdot\varphi(\tau_2x_2+\tau_3x_3+\mathbf{w_1}^T\boldsymbol{\tau})\cdot\{\beta_2x_2+\beta_3x_3+\mathbf{w_2}^T\boldsymbol{\beta}\} \\
&\quad -\Phi(\tau_2x_2+\tau_3x_3+\mathbf{w_1}^T\boldsymbol{\tau})\cdot\beta_3.
\end{aligned}
$$

Table 2: Estimation Results of 2. Stage OLS and the Two-Part Model (2PM) on Distance driven.

	Interaction Terms Included:				No Interaction Terms:			
	$\tau_{12} \neq 0, \tau_{13} \neq 0, \beta_{12} \neq 0, \beta_{13} \neq 0$				$\tau_{12} = \tau_{13} = \beta_{12} = \beta_{13} = 0$			
	2. Stage OLS		2PM		2. Stage OLS		2PM	
	Coeff.s	Errors	Effects	Errors	Coeff.s	Errors	Effects	Errors
female	**-2.641	(0.491)	* -0.224	(0.115)	**-1.491	(0.298)	** -0.581	(0.114)
employed	**-1.177	(0.422)	** 0.432	(0.130)	**-1.212	(0.413)	** 0.502	(0.127)
commute distance	** 0.030	(0.007)	0.005	(0.003)	** 0.029	(0.007)	0.003	(0.003)
age	**-0.034	(0.013)	0.005	(0.005)	* -0.033	(0.013)	0.005	(0.005)
female $\times$ *# children*	0.517	(0.294)	** 0.497	(0.179)	–	–	** -0.623	(0.133)
high-school diploma	** 0.981	(0.314)	** 0.432	(0.130)	** 1.024	(0.316)	** 0.502	(0.127)
# children	* -0.577	(0.235)	** 0.620	(0.064)	-0.277	(0.157)	** 0.722	(0.063)
# employed	-0.015	(0.240)	-0.056	(0.097)	-0.022	(0.240)	0.003	(0.095)
enoughcars	0.746	(0.422)	** 1.856	(0.108)	** 1.368	(0.292)	** 1.966	(0.107)
female $\times$ *enoughcars*	* 1.291	(0.552)	** 1.874	(0.208)	–	–	** -0.378	(0.034)
city region	* -0.713	(0.288)	-0.083	(0.116)	* -0.712	(0.289)	0.079	(0.114)
# observations used for estimation: 17,798								

Note: * denotes significance at the 5 %-level and ** at the 1 %-level, respectively. In the 2PM, interaction terms, such as *female* $\times$ *enoughcars*, stand for the interaction effect, here $\frac{\Delta^2 E}{\Delta x_2 \Delta x_1}$. Estimates of the marginal and interaction effects are average effects that have been computed by averaging individual effects over the observations.
The comparison between OLS and 2PM provided here should not be interpreted as the comparison between estimation results that disregard the censored data and those results that consider censoring. Rather, Table 2 compares incorrectly (2. stage OLS) and correctly (2PM) calculated marginal and interaction effects.

vanishing interaction effect of two variables indicates differing marginal effects of one variable conditional on alternative values of the other variable, as one would expect for two interacting variables. The concrete size of an interaction effect, however, hardly conveys any economic information. More easy to grasp are the conditional marginal effects pertaining to two variables that are assumed to interact.

In linear specifications, so-called interaction terms, consisting of the product $x_1 x_2$ of two explanatory variables, are typically included to capture the interaction effect, that is, the impact of an explanatory variable x_1 on the marginal effect of another explanatory variable x_2. In non-linear models, however, the marginal effect $\frac{\partial E}{\partial(x_1 x_2)}$ of the interaction term generally differs from the interaction effect, which is formally given by the second derivative $\frac{\partial^2 E}{\partial x_2 \partial x_1}$. This difference, along with the fact that interaction effects are generally non-vanishing even when no interaction terms are included in any non-linear specification, raises the question as to whether interaction terms are irrelevant in non-linear contexts.

It might be argued that it is not necessary to include any interaction term in non-linear specifications, such as the 2PM, as in this case the marginal effect of an explanatory variable x_1 generally depends on all other variables. This line of reasoning would be incorrect, however, since this dependence always prevails, irrespective of whether a particular effect of another variable x_2 is taken into account by including the interaction term $x_1 x_2$.

This can be seen from general expression (5), describing the marginal effect of variable x_1:

$$\frac{\partial E}{\partial x_1} = F'(u)(\beta_1 + \beta_{12} x_2).$$

The derivative $F'(u)$ captures the impact of a marginal change in $u = \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \mathbf{w}^T \boldsymbol{\beta}$ induced by the variation of any of the included variables, whereas a special effect of varying x_2 is only to be observed if an interaction term $x_1 x_2$ is included and the respective coefficient β_{12} is non-vanishing. In sum, the inclusion of interaction terms such as *female* $\times$ *age* is indispensable if one wants to meaningfully test, for example, the hypothesis of whether there are gender-specific differences in the impact of age on distance driven.

Appendix A: Data

Table A1: Variable Definitions and Descriptive Statistics

Variable Definition	Variable Name	Mean	Std. Dev.
Daily Kilometers driven for non-work purposes	*s*	4.505	10.801
Kilometers from home to work	*commute distance*	14.097	25.659
Dummy: 1 if person is female	*female*	0.480	–
Dummy: 1 if person is employed in a full-time or part-time job	*employed*	0.573	–
Age of the person	*age*	47.531	15.175
Dummy: 1 if person has a high school diploma	*high-school diploma*	0.340	–
Number of children younger than 18	*# children*	0.553	0.894
Number of employed household members	*# employed*	1.165	0.884
Dummy: 1 if number of cars $\geq$ number of licensed drivers	*enoughcars*	0.565	–
Dummy: 1 if household resides in a city	*city region*	0.323	0.468
Dummy: 1 if household has a private parking space	*private parking*	0.858	–
Walking time to the nearest public transportation stop	*minutes*	5.580	4.685
Dummy: 1 if the nearest public transportation stop is serviced by rail transit	*rail transit*	0.109	–

References

AI, C. , NORTON, E. C. (2003) Interaction terms in logit and probit models. *Economics Letters* **80**, 123-129.

DOW, W. H. , NORTON, E. C. (2003) Choosing Between and Interpreting the Heckit and Two-Part Models for Corner Solutions. *Health Services & Outcomes Research Methodology* **4**, 5-18.

DUAN, N. , MANNING, W. , MORRIS, C. , NEWHOUSE, J. (1984) Choosing between the Sample-Selection Model and the Multi-Part Model. *Journal of Business and Economic Statistics* **2**, 283-289.

STABUA (Federal Statistical Office) (2006) Germany's Population by 2050: Results of the 11th coordinated population projection. `http://www.destatis.de`

FRONDEL, M. , VANCE, C. (2010) Driving for Fun? Comparing the Effect of Fuel Prices on Weekday and Weekend Fuel Consumption. *Energy Economics*, **32 (1)**, 102-109.

FRONDEL, M. , VANCE, C. (2009) Do High Oil Prices Matter? Evidence on the Mobility Behavior of German Households. *Environmental and Resource Economics*, **43 (1)**, 81-94.

FRONDEL, M. , PETERS, J. , VANCE, C. (2008) Identifying the Rebound: Evidence from a German Household Panel. *The Energy Journal*, **29 (4)**, 154-163.

GREENE, W. H. (2010) Testing hypotheses about interaction terms in non-linear models. *Economics Letters* **107**, 291-296.

GREENE, W. H. (2007) *Limdep version 9.0, Econometric modeling guide.* Plainview, New York: Econometric Software Inc.

HECKMAN, J. (1979) Sample selection bias as a specification error. *Econometrica* **47**, 153-161.

KBA (Kraftfahrt Bundesamt) (2011) Stock of private cars. `http://www.kba.de/nn_268956/DE/Statistik/Fahrzeuge/Bestand/Halter/z__b__halter.html`

LEE, B. , MCDONALD, J. (2003) Determinants of commuting time and distance for Seoul residents: The impact of family status on the commuting of women. *Urban Studies* **40 (7)**, 1283-1302.

MCDONALD, N. (2005) Does the Built Environment Affect the Travel 'Gender Gap'? Conference Proceedings 35: Research on Women's Issues in Transportation; Volume 2: Technical Papers. Washington, DC: Transportation Research Board.

NORTON, E. C. , WANG, H. , AI, C. (2004) Computing Interaction Effects and standard errors in Logit and Probit Models. *The Stata Journal* **4 (2)**, 103-116.

MOP (2011) German Mobility Panel. `http://www.ifv.uni-karlsruhe.de/MOP.html`

VANCE, C. (2009) Marginal effects and significance testing with Heckman's sample selection model: A methodological note. *Applied Economics Letters* **16**, 1415-1419.

WHITE, M. (1986) Sex differences in urban commuting patterns. *American Economic Review* **76(2)**, 368-372.