

Huber, Peter; Oberhofer, Harald; Pfaffermayr, Michael

**Working Paper**

## Who creates jobs? Estimating job creation rates at the firm level

Working Papers in Economics and Finance, No. 2013-05

**Provided in Cooperation with:**

Department of Social Sciences and Economics, University of Salzburg

*Suggested Citation:* Huber, Peter; Oberhofer, Harald; Pfaffermayr, Michael (2013) : Who creates jobs? Estimating job creation rates at the firm level, Working Papers in Economics and Finance, No. 2013-05, University of Salzburg, Department of Social Sciences and Economics, Salzburg

This Version is available at:

<https://hdl.handle.net/10419/100637>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

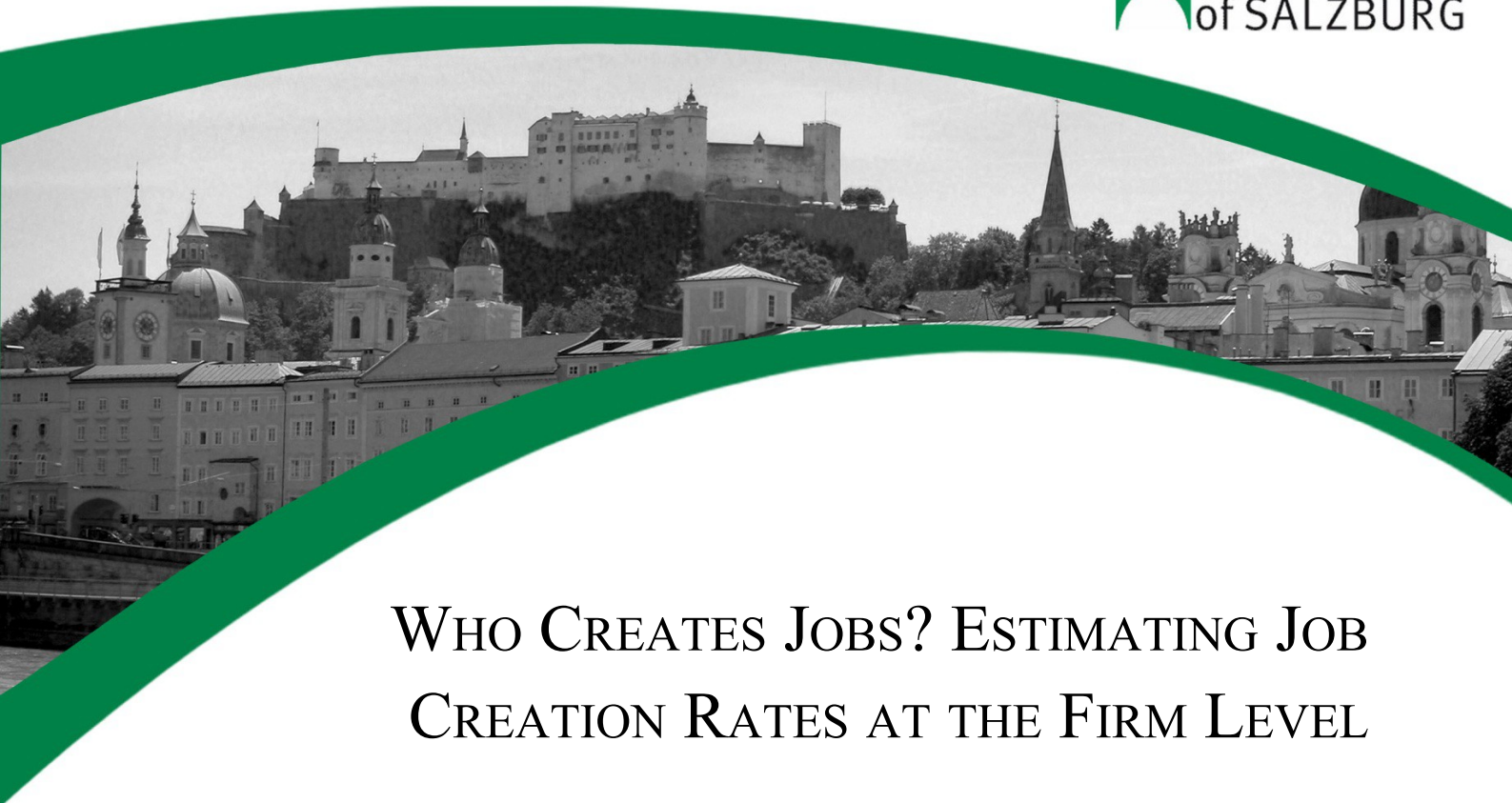
Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



# WHO CREATES JOBS? ESTIMATING JOB CREATION RATES AT THE FIRM LEVEL

PETER HUBER, HARALD OBERHOFER AND MICHAEL PFAFFERMAYR

WORKING PAPER No. 2013-05



**WORKING PAPERS** IN  
ECONOMICS AND FINANCE

# Who Creates Jobs?

## Estimating Job Creation Rates at the Firm Level\*

Peter Huber\*, Harald Oberhofer\*\*, Michael Pfaffermayr\*\*\*

### Abstract

This paper analyzes econometric models of the Davis, Haltiwanger and Schuh (1996) job creation rate. In line with the most recent job creation literature, we focus on employment-weighted OLS estimation. Our main theoretical result reveals that employment-weighted OLS estimation of DHS job creation rate models provides biased marginal effects estimates. The reason for this is that by definition, the error terms for entering and exiting firms are non-stochastic and non-zero. This violates the crucial mean independence assumption requiring that the conditional expectation of the errors is zero for all firms. Consequently, we argue that firm entries and exits should be analyzed with separate econometric models and propose alternative maximum likelihood estimators which are easy to implement. A small-scale Monte Carlo analysis and an empirical exercise using the population of Austrian firms point to the relevance of our analytical findings.

**Keywords:** DHS job creation rate; firm size; firm age; maximum likelihood estimation; three-part model; multi-part model; Monte Carlo simulation.

**JEL:** C18; C53; D22; E24; L25; L26; M13.

---

\*We are grateful to John Haltiwanger, Ron Jarmin and Javier Miranda for their excellent comments on an earlier version of this paper and valuable discussions. We would also like to thank the participants of the 5<sup>th</sup> Workshop on Empirical Industrial Organization 2012 at the Austrian Institute of Economic Research in Vienna, of the Workshop on Firm Growth and Innovation 2012 in Tarragona, of the 9<sup>th</sup> meeting of the European Network on the Economics of the Firm (ENEF) 2012 in Bologna, of the annual meeting of the Austrian Economic Association 2013 in Innsbruck, the Applied Economics meeting 2013 in Granada and seminar participants at the Universities of Bayreuth, Graz, Innsbruck, Linz and the Vienna University for Economics and Business for their comments and suggestions. Financial support from the ‘Oesterreichische Nationalbank’ (OeNB, grant numbers 13370 and 14383) is gratefully acknowledged.

\*Austrian Institute of Economic Research, Arsenal, Objekt 20, A-1030 Vienna, Austria. Email: huber@wifo.ac.at.

\*\*Department of Economics and Social Sciences and Salzburg Centre of European Union Studies (SCEUS), University of Salzburg, Residenzplatz 9, 5010 Salzburg, Austria, The Austrian Center for Labor Economics and the Analysis of the Welfare State. E-mail address: harald.oberhofer@sbg.ac.at.

\*\*\*Department of Economics, University of Innsbruck, Universitaetsstr. 15, A-6020 Innsbruck, Austria, Austrian Institute of Economic Research and CESifo, E-mail: michael.pfaffermayr@uibk.ac.at.

# 1 Introduction

During the last three decades the question on which firms are the most important net job creators has triggered heated discussions in both the academic community and among policy makers. Starting with the early insights provided by Birch (1979) the debate centered around the issue whether small or large firms are more successful in creating jobs. A recent study by Neumark *et al.* (2011) reinforced the crucial role of small firms for net job creation while Haltiwanger *et al.* (2013a), applying employment-weighted OLS estimation, highlighted the so far neglected role of firm age. Based on an impressive sample of US firms they find that young firms, irrespective of their size, are the most important contributors to net-job creation.

A careful empirical analysis of job creation across different types of firms (small, large, young and old ones) has to take into account several different sources of net job creation or destruction. By definition, newly founded firms create jobs while exiting firms destroy them. Additionally, continuing firms adjust their size and might either increase or decrease their level of employment. Taking these arguments together, when analyzing the determinants of net job creation, one has to simultaneously examine firm entry, firm exit, firm contractions and firm growth to address this issue accurately (see, e.g., Spletzer 2000). In their seminal contributions, Davis and Haltiwanger (1992) and Davis, Haltiwanger and Schuh (1996) proposed a new measure of job creation that permits such an integrated treatment of firm exit, firm entry and employment changes in continuing firms. In particular, the Davis-Haltiwanger-Schuh (DHS) measure (as discussed in more detail below) is defined in a way that the job creation rate of exiting (entering) firms amounts to -2 (2) while continuing firms exhibit job creation rates in the open interval between -2 and 2. This job creation measure provides a convenient way to (descriptively) calculate the relative importance of firm entry and firm exit for net job creation.

So far, several alternative methodological approaches have been put forward in order to empirically address the question on which firms are most important (overall) net job creators. Due to the limited availability of data capturing the whole population of firms and/or plants, the historically first strand of the literature was limited to analyzing this question at a more aggregated level. Consequently, early contributions in this literature constructed average DHS job creation rates at e.g., age-size-industry-year cells and used the respective cell-means as observational unit in a simple regression framework estimated via (weighted) OLS (see, e.g., Dunne *et al.* 1989; Baldwin *et al.* 1998; Faberman 2003; Stiglbauer *et al.* 2003 and Armington and Acs 2004). More recent contributions include

Moscarini and Postel-Vinay (2012) and Haltiwanger *et al.* (2013a) where the latter also show that this cell-means approach is equivalent to applying employment-weighted OLS to firm level data.<sup>1</sup> Second, some authors argue that the DHS job creation rate induces censoring at the  $[-2, 2]$  interval and, thus, apply Tobit estimation to firm and/or plant level data (see, e.g., Ibsen and Westergaard-Nielsen 2005; Ilmakunnas and Maliranta 2005 and Guertzgen 2009). The Tobit model, however, seems to be inappropriate in this case, since it assumes a theoretical distribution of firm level job creation rates with support outside the  $[-2, 2]$  interval, while the DHS job creation rate, by definition, is bounded by -2 and 2. Finally, the most recently established methodological approach also relies on firm and/or plant level data and applies (weighted) OLS estimation to one- and two-way models (see, e.g., Burgess *et al.* 2000; Haltiwanger and Vodopivec 2003; Voulgaris *et al.* 2005) or to saturated models containing firm size and age dummies as well as interactions thereof (see, e.g., Haltiwanger *et al.* 2013a).

This paper aims at contributing to the job creation literature by analyzing econometric models of the DHS job creation rate. We will specifically focus on the third strand of the empirical literature which applies employment-weighted OLS estimation to a saturated model. Since, this approach is equivalent to constructing cell-means as the unit of observation, our findings are also applicable to the earlier literature. Our main result reveals that, in general, employment-weighted OLS estimation of the DHS job creation delivers inconsistent and biased parameter and marginal effects estimates. The reason for this is that by definition, the error terms for entering and exiting firms are non-stochastic and non-zero. Even in saturated models this violates the crucial mean independence assumption requiring that the conditional expectation of the errors is zero for all firms. We further illustrate the bias in a saturated two-way interaction model and, thereby, are able to show that the bias increases with the (absolute) net-entry rate (i.e., entry minus exit) and it disappears if each age-size cell either contains exclusively entering, continuing or exiting firms. Therefore, pooling of entering, exiting and continuing firms in single cells necessarily yields biased and inconsistent estimates.

With this finding at hand and in line with the empirical firm growth literature (see, e.g., Sutton 1997; Hart 2000 and Coad 2009 for surveys), we propose an alternative three-part estimation procedure that allows to obtain unbiased estimators and is simple to implement. Based on the assumption of a lognormal firm size distribution for the continuing firms we formulate a maximum-likelihood estimator that is equivalent to applying simple OLS estimation for employment growth measured in terms of the log difference for con-

---

<sup>1</sup>Actually, the parameters of the weighted OLS regression coincide with the cell means in a saturated model that includes all interaction effects as shown by, e.g. Searle (1987, p. 102).

tinuing firms only. This together with a probit model for exiting firms and accounting for job creation by entering firms constitute our three-part approach, which does not need to pool over all observations as it is the case in the saturated and employment-weighted model for the DHS job creation rate. We also briefly discuss a possible generalization to a multi-part model which allows the marginal job creation effects to differ between growing and declining (continuing) firms and is less restrictive in comparison to the lognormal three-part model.

In a next step we validate our theoretical findings with both, a small-scale Monte Carlo analysis and an empirical application for the population of Austrian firms observed over the period 1993 to 2009. The Monte Carlo analysis confirms our theoretical results documenting biased point estimates in the employment-weighted OLS regressions that pool over all observations. By contrast, the alternative three-part model delivers unbiased estimates. Our empirical exercise reveals (substantial) differences in the (employment-weighted) estimated marginal firm size and age effects across the different econometric models, again lending support to our theoretical finding. The bias from employment-weighted OLS estimation is most pronounced for the groups of smallest and youngest firms. With regard to firm age we are able to confirm that all employment-weighted OLS estimates are downward biased, underestimating the (true) role of firm age. With regard to firm size, the employment-weighted OLS estimator underestimates the (employment-weighted) job creation rates for the smallest Austrian firms.

## 2 The econometrics of job creation rates

The three different sources for net job creation at the firm-level are entry, employment changes in continuing firms and exit. Formally, in each year  $t$  the population of firms can be partitioned into three groups (see Horrace and Oaxaca 2006 for a similar approach in the context of the linear probability model).

$$\begin{aligned}
 G_x &= \{i|y_{it} = 0\} \text{ (exiting firms)} \\
 G_c &= \{i|y_{i,t-1} \neq 0 \text{ and } y_{it} \neq 0\} \text{ (continuing firms)} \\
 G_n &= \{i|y_{i,t-1} = 0\} \text{ (entrants),}
 \end{aligned}$$

with corresponding probabilities  $p_k = P(i \in G_k)$ ,  $k = x, c, n$  with  $p_x + p_c + p_n = 1$ . In empirical applications, these probabilities may be specified as a function of exogenous explanatory variables and thus may be firm-specific.

Following the job creation literature and in line with Davis and Haltiwanger (1992) and Davis, Haltiwanger and Schuh (1996), we measure a firm  $i$ 's net job creation over the period  $t - 1$  to  $t$ , by

$$g_{it} = 2 \frac{y_{it} - y_{i,t-1}}{y_{it} + y_{i,t-1}} = \begin{cases} -2 & \text{if } i \in G_x \\ 2 \frac{y_{it}/y_{i,t-1}-1}{y_{it}/y_{i,t-1}+1} & \text{if } i \in G_c \\ 2 & \text{if } i \in G_n, \end{cases}$$

where  $y_{it}$  denotes a firm's number of employees at time  $t$ . The main advantage of this measure is that the (net) job creation rate is defined for *all* observations, i.e., also for entrants ( $y_{i,t-1} = 0$ ) and exiting firms ( $y_{it} = 0$ ) and it can easily be used to calculate aggregate net job creation rates, e.g., for size or age groups and/or for specific industries. However, this convenience comes at the cost of spikes (i.e., non-linearities) of the distribution of  $g_{it}$  at  $-2$  and  $2$  in the presence of positive entry and exit rates (probabilities). In a list of 10 alternative measures surveyed by Tornqvist *et al.* (1985),  $g_{it}$  (denoted there as  $H_3$ ) is shown to be a useful measure for relative changes, but to be non-additive in the time dimension. The log difference  $\ln(y_{it}/y_{i,t-1})$  is found to be preferable as it is the only measure of relative change that is symmetric, additive and normed. The drawback of the log differences as a measure of relative change, however, is that it is not defined for exiting and entering firms with  $y_{it} = 0$  and  $y_{i,t-1} = 0$ , respectively.

Econometric models for DHS job creation rates typically pool over all three groups of entering, exiting and continuing firms and, hence, use *all* observations. The resulting model may be written as

$$g_{it} = \mathbf{x}'_{it}\beta + \varepsilon_{it}, \tag{1}$$

where the set of exogenous covariates are collected in a  $(K \times 1)$  vector  $\mathbf{x}_{it}$  with the corresponding parameter vector  $\beta$ .  $\varepsilon_{it}$  denotes the *iid* error term. However, this approach suffers from drawbacks both from an economic and an econometric point of view.

In economic terms, the pooling assumption leads to rather implausible comparative statistics. A change in any covariate is restricted to affect the expected DHS job creation rate of continuing firms in the same way as that of exiting and entering firms. In view of the empirical evidence provided by the firm growth literature and the IO literature on the determinants of market exit (see, e.g., Caves 1998), this assumption lacks empirical support. To give an example, the pooling assumption requires that the expected DHS job creation rate is the same for both, continuing and exiting small firms. Jovanovic's (1982) learning theory, by contrast, states that small (and young) firms, which fail to adjust their firm size to an efficient level rapidly, will be more likely to leave the market while

surviving small (and young) firms tend to grow faster, on average.

Econometrically, it can be shown (see the Appendix) that applying weighted OLS to eq. (1) (see, e.g., Burgess *et al.* 2000; Haltiwanger and Vodopivec 2003; Voulgaris *et al.* 2005 and Haltiwanger *et al.* 2013a) yields biased and inconsistent estimates. In general, the bias results from pooling of exiting, entering and continuing firms in a single model and the lack of variation in the error terms for exiting and entering firms. These firms have either  $g_{it} = 2$  or  $g_{it} = -2$  and, therefore, non-stochastic errors given by  $\varepsilon_{it} = -2 - \mathbf{x}'_{it}\beta$  and  $\varepsilon_{it} = 2 - \mathbf{x}'_{it}\beta$ , respectively. As a consequence the mean independence assumption for the validity of OLS, namely that  $E[\varepsilon_{it}|\mathbf{x}_{it}] = 0$ , is violated for exiting and entering firms, since  $E[\varepsilon_{it}|\mathbf{x}_{it}, d_{i,x} = 1] = -2 - \mathbf{x}'_{it,x}\beta$  and  $E[\varepsilon_{it}|\mathbf{x}_{it}, d_{i,n} = 1] = 2 - \mathbf{x}'_{it,n}\beta$ , where the indices  $x$  and  $n$  refer to exiting and entering firms and  $d_{i,x}$ ,  $d_{i,c}$  and  $d_{i,n}$  are indicator variables for exiting, continuing and entering firms, respectively.

In many empirical applications of  $g_{it}$  as dependent variable, the explanatory variables include dummies for size and age classes, in addition to dummies for industries and time. In general and under OLS assumptions a fully saturated model that includes dummies for age and size classes as well as interactions thereof obtains the respective cell means as unbiased estimates regardless of the distribution of disturbances (see Angrist and Pischke 2009, pp. 37 and 48-51). However, this proposition requires that the mean independence assumption  $E[\varepsilon_{it}|\mathbf{x}_{it}] = 0$  holds for all  $i$ , where in this example  $\mathbf{x}_{it}$  includes the full set of dummy variables that define the saturated model. Since this condition is violated due to the presence of entering and exiting firms, the bias can only be avoided by specifying separate models for these entry and exit cells, respectively. As discussed above, this is also more plausible from an economic point of view.

As a further consequence, econometric models that ignore this issue, pool firms in age-size-industry-year cells and use the respective weighted cell-means as observational unit in a regression framework are also prone to bias (see, e.g., the applications in Dunne *et al.* 1989 and Moscarini and Postel-Vinay 2012). The corresponding cell-means are biased predictors and as a result (weighted) OLS estimates based on these cell means are also biased.

In order to provide a simple illustration for the bias of the employment-weighted OLS-estimator in models with  $g_{it}$  as dependent variable consider a saturated two-way interaction model that only contains dummies for size and age groups and interactions thereof. The first age group refers to entering firms with zero age, which can be interpreted as the inclusion of an entry dummy variable. We index the corresponding cells by  $h = 1, \dots, H$ ,



(see, e.g., Searle, 1987, chapter 4 for a discussion of such a model). For simplicity, we consider an age-size cell  $h$  in a cross-section of DHS job creation rates,  $g_{ih}, i = 1, \dots, n_h$ , between two periods  $t$  and  $t - 1$  and skip the time index. The corresponding econometric model is thus given by  $g_{ih} = \mu_h + \varepsilon_{ih}$  and the cell weights are denoted by  $w_{ih}$  with  $\sum_{i=1}^{n_h} w_{ih} = 1$  and are typically specified as  $w_{ih} = \frac{y_{ih} + y_{ih,-1}}{\sum_{k=1}^{n_h} y_{kh} + y_{kh,-1}}$ . Following the literature (see, e.g. Haltiwanger *et al.* 2013a), we treat these weights as fixed and exogenously given.<sup>2</sup> The weighted OLS estimator minimizes the weighted sum of squared residuals ( $WS$ ). Thus, the first order condition and the corresponding weighted OLS estimator read as

$$WS = \sum_{h=1}^H \sum_{i=1}^{n_h} w_{ih} (g_{ih} - \mu_h)^2$$

$$\frac{\partial WS}{\partial \mu_h} = -2 \sum_{i=1}^{n_h} w_{ih} (g_{ih} - \mu_h) \Rightarrow \hat{\mu}_h = \sum_{i=1}^{n_h} w_{ih} g_{ih}.$$

$\hat{\mu}_h$  can be rewritten as

$$\begin{aligned} \hat{\mu}_h &= \sum_{i=1}^{n_h} w_{ih} g_{ih} d_{ih,x} + \sum_{i=1}^{n_h} w_{ih} g_{ih} d_{ih,c} + \sum_{i=1}^{n_h} w_{ih} g_{ih} d_{ih,n} \\ &= - \sum_{i=1}^{n_h} w_{ih} 2d_{ih,x} + \sum_{i=1}^{n_h} w_{ih} (\mu_j + \varepsilon_{ih}) d_{ih,c} + \sum_{i=1}^{n_h} w_{ih} 2d_{ih,n}. \end{aligned}$$

Using  $E[\varepsilon_{ih} | d_{ih,c} = 1] = 0$  and  $\sum_{i=1}^{n_h} w_{ih} = 1$ , the bias of  $\hat{\mu}_h$  can then be derived as<sup>3</sup>

$$\begin{aligned} E[\hat{\mu}_h] &= E \left[ - \sum_{i=1}^{n_h} w_{ih} 2d_{ih,x} + \sum_{i=1}^{n_h} w_{ih} (\mu_j + \varepsilon_{ih}) d_{ih,c} + \sum_{i=1}^{n_h} w_{ih} 2d_{ih,n} \right] \\ &= 2(p_{h,n} - p_{h,x}) + (1 - p_{h,x} - p_{h,n})\mu_h \\ E[\hat{\mu}_h - \mu_h] &= -(p_{h,x} + p_{h,n})\mu_h + 2(p_{h,n} - p_{h,x}). \end{aligned} \tag{2}$$

---

<sup>2</sup>It can be shown that the typically used weights are endogenous. Consequently, the weighted OLS estimator of the saturated model is a highly non-linear function of  $g_{ih}$ . To our knowledge, this issue has so far been neglected in the literature on the estimation of DHS job creation rates and the econometric properties of this estimator still have to be established. This, however, is beyond the scope of our paper, but we take up this issue in our Monte Carlo experiments discussed below.

<sup>3</sup>With firm-specific state-probabilities, the bias is given as

$$E[\hat{\mu}_h - \mu_h] = 2 \left( \sum_{i=1}^{n_h} w_{ih} p_{ih,n} - \sum_{i=1}^{n_h} w_{ih} p_{ih,x} \right) - \sum_{i=1}^{n_h} w_{ih} \mu_h (p_{ih,n} + p_{ih,x}).$$

With exogenously given weights, the bias is dominated by  $-(p_{h,x} + p_{h,n})\mu_h$  if the difference in the entry and exit probabilities is small. This would be the case if expected net-entry rates are low (i.e.,  $p_{h,x} \approx -p_{h,n}$ ). Moreover, the bias will be negative and more pronounced the larger the sum of entry and exit probabilities are or, equivalently, the lower the probability weight of the continuing firms is. The sign of the bias is unambiguously negative if  $p_{h,x} > p_{h,n}$ , (i.e., the probability weight of exiting firms is larger than that of entering firms). This result also illustrates that we obtain unbiased estimates if each age-size cell either contains exclusively entering, continuing or exiting firms, but no combination of them. In this case we have either  $p_{h,n} = 1$  or  $p_{h,x} = 1$  or  $p_{h,c} = 1$ .

To illustrate the connection to the general model derived in the Appendix, we denote the matrix of the weighted explanatory variables by  $\mathbf{X}^*$  and  $\mathbf{e}^*$  a vector of ones multiplied elementwise with the weights  $w_{it}$ , respectively. The bias then amounts to

$$E[\widehat{\beta} - \beta | \mathbf{X}^*] = -(p_x + p_n)\beta + 2(p_n - p_x) (\mathbf{X}^{*\prime} \mathbf{X}^*)^{-1} \mathbf{X}^{*\prime} \mathbf{e}^* \quad (3)$$

In the saturated model discussed above each column in  $\mathbf{X}^*$  refers to a dummy for one cell multiplied elementwise with the corresponding weights. It can easily be verified that in this case  $(\mathbf{X}^{*\prime} \mathbf{X}^*)^{-1} \mathbf{X}^{*\prime} \mathbf{e}^*$  reduces to a vector of ones so that (3) reverts to (2). Again, it becomes clear that the bias originates from pooling all three types of firms into one single model.

In the following we present two alternative procedures, which allow to consistently estimate the DHS job creation rate and are easy to implement. To start with, one can estimate separate equations for the entering, exiting and continuing firms in a three-part model. In particular, for continuing firms we consider the log change in firm size  $l_{it} = \ln(y_{it}) - \ln(y_{i,t-1})$  and derive the implied employment growth rate as

$$g(l_{it}) = 2 \frac{e^{l_{it}} - 1}{e^{l_{it}} + 1}$$

$$l_{it} = \ln z_{it} + \ln \eta_{it},$$

where  $\ln z_{it}$  denotes the conditional mean.<sup>4</sup>  $\ln \eta_{it}$  is an *iid* random disturbance with expectation 0. For continuing firms the conditional expectation of  $g_{it}$  is non-linear in  $l_{it}$

---

<sup>4</sup>Alternatively, one could consider the transformation  $g_{it} \rightarrow l(g_{it})$ . This means specifying the true model as  $g_{it} = z_{it} + \eta_{it}$  and deriving the implied linear approximation  $l(g_{it})$ . For empirical applications, this alternative transformation is not useful, since it does not guarantee that  $-2 \leq g_{it} \leq 2$  and, furthermore, in this case  $l_{it}$  might not be defined for continuing firms. In general, the predicted values of  $g_t$  will deviate from 2 (-2) in case of entry (exit) when assuming this alternative data generating process.

and the linear approximation of  $g(l_{it})$  at  $l_{it} = 0$  is given by

$$g(l_{it}) = 2 \frac{e^{l_{it}} - 1}{e^{l_{it}} + 1} \approx 0 + 2 \frac{e^{l_{it}} + 1 - (e^{l_{it}} - 1)}{(e^{l_{it}} + 1)^2} e^{l_{it}} \Big|_{l_{it}=0} \quad l_{it} = 2 \frac{2}{(2)^2} l_{it} = l_{it},$$

implying that for continuing firms a regression of  $g_{it}$  on a set of explanatory variables in logs may be interpreted as a linear approximation of  $g_{it}$  in terms of  $l_{it}$ .

For simplicity, for the group of continuing firms we assume that the error term  $\eta_{it}$  is distributed as *iid* lognormal so that a consistent ML estimator can easily be derived. However, other distributional assumptions are possible. For example, to account for the non-(log-)normal and asymmetrically skewed distribution of firm growth rates among continuing firms one could use the asymmetric exponential power distribution as proposed by e.g., Komunjer (2007), Zhu and Zinde-Walsh (2009) and Bottazzi and Secchi (2011). In general the asymmetric exponential power distribution is able to deal with ‘tent-shaped’ type of empirical distributions with asymmetric tails.<sup>5</sup> Alternatively, a multi-part model based on five types of firms (exiting, shrinking, stagnating, expanding and entering firms, respectively) would explicitly account for a mass of zeros and asymmetric size adjustments.

The growth rate distribution of continuing firms (i.e., conditional on  $y_{it} \neq 0$  and  $y_{i,t-1} \neq$

---

Analytically, for  $-2 < g_{it} < 2$  we have

$$l(g_{it}) = \ln(2 + g_{it}) - \ln(2 - g_{it}) = \ln(2 + z_{it} + \eta_{it}) - \ln(2 - z_{it} - \eta_{it}), \quad \text{and}$$

$$l(g_{it}) = \begin{cases} -\infty & \text{if } g_{it} = -2 \text{ (exit)} \\ \ln(2 + g_{it}) - \ln(2 - g_{it}) & \\ \infty & \text{if } g_{it} = 2 \text{ (entry)} \end{cases}$$

This model implies that the conditional expectation of  $g_{it}$  is linear, while that of  $l_{it}$  is non-linear. However, at  $-2 < g_{it} < 2$  the linear approximation of  $l_{it}$  at 0 yields  $l_{it} \approx g_{it}$ , since

$$\ln(2 + g_{it}) - \ln(2 - g_{it}) \approx \ln(2) - \ln(2) + \frac{1}{2}g_{it} - \frac{1}{2}(-1)g_{it} = g_{it}.$$

<sup>5</sup>Zhu and Zinde-Walsh (2009) show that the asymptotic properties of an asymmetric exponential power distribution based maximum likelihood estimator are only valid when both tail parameters exceed 0.5. For the data at hand (which will be discussed in Section 4) this criterion is not fulfilled and, therefore, the asymmetric exponential power distribution does not form a valuable alternative for our case. One potential reason for this might be that more than 40 percent of all observed job creation rates exactly amount to 0.

0), measured in log differences, under the assumption of log-normality is given by<sup>6</sup>

$$f(l_{it}|y_{it} \neq 0, y_{i,t-1} \neq 0) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-\frac{1}{2} \frac{(l_{it} - \ln(z_{it}))^2}{\sigma^2}}. \quad (4)$$

Then, in a three-part model the contribution to the likelihood function referring to exiting and continuing firms in period  $t$  can be written as

$$L_t(\gamma, \beta, \sigma) = \prod_{G_x} p_{it,x}(\gamma; x_{it}) \prod_{G_c} (1 - p_{it,x}(\gamma; x_{it})) f(\beta, \sigma; l_{it}, x_{it}).$$

In this model the entry probability is not included because firm-specific pre-entry characteristics are not observable. Consequently, in order to incorporate the third part of the model,  $p_{it,n}$ , we add the employment-weighted job creation rates for entering firms within each cell. The parameter vector of the model for the probability of exit  $p_{it,x}$  may be estimated by a separate Probit model. The parameters of the specification for the continuing firms,  $\beta$  and  $\sigma^2$ , can be estimated by maximizing the likelihood based on the density from equation (4) excluding the observations referring to entering and exiting firms. As demonstrated in Footnote 6, the ML estimates of the  $\beta$  parameters are numerically equivalent to the OLS estimates in a regression with the left hand side variable measured as log difference in firm size.

A more general multipart-model may be motivated by the literature on employment adjustment under non-convex adjustment costs (see, e.g., Di Iorio and Fachin 2004, Nilsen et al. 2007 and Lapatinas 2009). This work suggests separate models for exiting, shrinking, stagnating and expanding firms and for entrants. Following e.g., Nilsen et al. (2007), one first estimates a general discrete choice model, in our case the multivariate probit (see Capellari and Jenkins 2003), based on the latent variable

$$y_{it,k}^* = \mathbf{x}_{it}' \gamma_k + \varepsilon_{it,k}, \quad k \in \{\text{exiting, shrinking, stagnating, expanding}\},$$

where  $\varepsilon_{it,k}$  is allowed to be correlated across the alternatives. The model is estimated by simulated maximum likelihood using the Geweke-Hajivassiliou-Keane smooth recursive

---

<sup>6</sup>To obtain the density in terms of  $g_{it}$  we use the transformation  $\eta_{it} = \frac{1}{z_{it}} \left( \frac{2+g_{it}}{2-g_{it}} \right)$  and  $\frac{\partial \eta_{it}}{\partial g_{it}} = \frac{1}{z_{it}} \frac{4}{(2-g_{it})^2}$ .

$$f(\eta(g_{it}) | -2 < g_{it} < 2) = (1 - p_{it,n} - p_{it,x}) \frac{1}{\sqrt{2\pi}} \frac{4}{(2+g_{it})(2-g_{it})} \frac{1}{\sigma} e^{-\frac{1}{2} \frac{(\ln(2+g_{it}) - \ln(2-g_{it}) - \ln(z_{it}))^2}{\sigma^2}}$$

Note that  $\ln(2+g_{it}) - \ln(2-g_{it}) = l_{it}$  and, thus, the parameter estimates of the resulting model are numerically identical to the OLS estimates of a model with  $l_{it}$  as the dependent variable.

conditioning simulator. This simulator is based on a useful property of the multivariate normal distribution function, namely that it can be decomposed in a product of sequentially conditioned univariate normal distribution functions that are easy to evaluate accurately.

The second part of the model postulates separate linear models for shrinking and growing firms using  $l_{it}$  (i.e., the log-difference in the number of employees) as dependent variable. These models can be estimated by OLS as shown for the more restrictive three-part model discussed above. Note, also in the multi-part model entering firms have to be treated as in the three-part model.

In the multi-part setting, model predictions are based on estimated probabilities from the multivariate normal and the two outcome equations for shrinking and growing firms using

$$\begin{aligned}\widehat{g}_{it,s} &= 2 \frac{e^{\widehat{l}_{it,s}-1}}{e^{\widehat{l}_{it,s}+1}} \\ \widehat{g}_{it,g} &= 2 \frac{e^{\widehat{l}_{it,g}-1}}{e^{\widehat{l}_{it,g}+1}} \\ \widehat{g}_{it,c} &= \widehat{p}_{it,s} \widehat{g}_{it,s} + \widehat{p}_{it,g} \widehat{g}_{it,g},\end{aligned}$$

where the index  $s(g)$  refers to shrinking (growing) firms. In the three-part model the corresponding predicted growth rate of exiting and continuing firms is calculated as

$$\widehat{g}_{it} = -2\widehat{p}_{it,x} + (1 - \widehat{p}_{it,x})\widehat{g}_{it,c}.$$

Lastly, and in line with the discussion from above we set the DHS job creation rate of entering firms to 2 as we do not model the industry-time specific entry rates explicitly.

In order to analyze the overall impact of firm size and firm age on net job creation, we have to aggregate individual job creation rates within different firm groups. Based on the DHS approach the job creation rates for various groups of firms (e.g., industry, size and age classes) with index  $h$  that are populated by  $n_t^h$  firms are calculated as (see also Haltiwanger *et al.* 2013b)

$$g_t^h = \sum_{i=1}^{n_t^h} \frac{y_{it} + y_{i,t-1}}{\sum_{k=1}^{n_t^h} y_{kt} + y_{k,t-1}} g_{it} = 2 \frac{\sum_{i=1}^{n_t^h} (y_{it} - y_{i,t-1})}{\sum_{k=1}^{n_t^h} y_{kt} + y_{k,t-1}} \quad (5)$$

or, equivalently, as

$$\begin{aligned}
g_t^h &= 2 \frac{\sum_{i=1}^{n_t^h - n_{x,t}^h - n_{n,t}^h} y_{it} - y_{i,t-1}}{\sum_{k=1}^{n_t^h} y_{kt} + y_{k,t-1}} + 2 \frac{\sum_{i=1}^{n_{t,n}^h} y_{it}}{\sum_{k=1}^{n_t^h} y_{kt} + y_{k,t-1}} \\
&\quad - 2 \frac{\sum_{i=1}^{n_{t,x}^h} y_{i,t-1}}{\sum_{k=1}^{n_t^h} y_{kt} + y_{k,t-1}} \\
&:= g_{t,c}^h + g_{t,n}^h + g_{t,x}^h.
\end{aligned} \tag{6}$$

Equations (5) and (6) show that there is no need to estimate a pooled model as in Haltiwanger *et al.* (2013a, see column 5 in their Table 2). By contrast, it is possible to recover the net job creation rates by calculating  $g_{t,c}^h$  or its predictions for the continuing firms and adding the corresponding (weighted) rates referring to the entering and exiting firms, respectively. Accordingly, the proposed three-part model is very simple to implement, since it only requires to estimate a probit model for exit together with an OLS estimation for employment growth measured in terms of the log difference for the continuing firms. Finally, one simply has to add the cell-weighted job creation rates of entering firms as demonstrated in the second term of equation (6). In a similar vein, the multi-part model only requires the estimation of a multivariate probit model together with two OLS estimators for the log-difference in employment for shrinking and expanding continuing firms, respectively.

### 3 Monte Carlo simulation

In order to analyze the properties of the above discussed estimators in finite samples, we generate DHS job creation rates between time 0 and  $T$  in a cross-section of continuing firms according to the following saturated econometric model:

$$\ln(y_{iT}) - \ln(y_{i0}) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \beta_4 d_{i,1} + \beta_5 d_{i,2} + \ln \eta_i, \tag{7}$$

where  $x_{i1}$  (firm size) takes the value of 1, if a generated lognormal random variable is smaller than its median value and  $x_{i2}$  (firm age) is a dummy which equals 1 if a generated uniform random variable is smaller than its respective median. Accordingly,  $x_{i1}$  and  $x_{i2}$  are equal to 1 for small and young firms, respectively and both variables are fixed in repeated samples.  $d_{i,1}$  and  $d_{i,2}$  are dummy variables for small and large entering firms, respectively, with the parameters  $\beta_4$  and  $\beta_5$  amounting to +2 by construction.  $\beta_0$  to  $\beta_3$

denote the (marginal effects) parameters to be estimated where  $\beta_0$  measures the average marginal job creation rate for the omitted reference group of large-old firms.

Log initial size,  $\ln(y_{i0})$ , is generated as *iid*  $N(0, 1)$  and also kept fixed in repeated samples. It is independent of  $\eta_i$ , so by construction this setting rules out any bias due to regression to the mean effects.<sup>7</sup>  $\ln \eta_i$  is generated as *iid*  $N(0, \sigma^2)$ . The true data generating process assumes parameter values  $\beta_0 = -0.1$ ,  $\beta_1 = \beta_2 = 0.1$  and  $\beta_3 = 0.05$ . In economic terms, these parameters reflect the main findings from the empirical firm growth literature, namely that smaller and younger surviving firms tend to grow faster.

In order to obtain a group of entering firms, we generate a Bernoulli random variable  $d_{i,n}$ , which sets  $y_{i0} = 0$  with probability  $p_n$ . The group of exiting firms is constructed in a way that smaller and younger firms are less likely to survive. More specifically, a firm exits from the market if a generated latent standard normal random variable is lower than  $-0.84 + 0.1(x_{i1} - 0.5) + 0.01(x_{i2} - 0.5) - 0.01(x_{i1}x_{i2} - 0.25)$ , guaranteeing an expected (aggregated) exit rate of  $p_{i,x} = 0.2$ . For young and small firms (i.e.,  $x_{i1} = 1$  and  $x_{i2} = 1$ ) this process imposes an exit probability of 0.41 while the large and old firms (i.e.,  $x_{i1} = 0$  and  $x_{i2} = 0$ ) exit with a probability of only 0.09. The just described exit For exiting firms we set  $y_{iT} = 0$ . Lastly, we calculate the DHS job creation rate which is given by  $g_{iT} = 2 \frac{y_{iT} - y_{i0}}{y_{iT} + y_{i0}}$ .

In the Monte Carlo analysis, we carry out alternative experiments that consider combinations of  $p_n \in \{0, 0.2\}$ ,  $p_{i,x} \in \{0, 0.2\}$  and  $\sigma \in \{0.1, 0.5\}$  and compare the saturated employment-weighted OLS estimator to the alternative maximum likelihood approach. The latter is equivalent to an unweighted OLS estimator for the log difference of continuing firms only. In Table 1 we report the bias, the root mean square error (RMSE) (both multiplied by 100) and the rejection rates under  $H_0$  of  $t$ -tests for  $\beta_0 = -0.1$ ,  $\beta_1 = \beta_2 = 0.1$  and  $\beta_3 = 0.05$ . The bias is calculated as the average deviation of the estimated parameters from their respective true values, while the RMSE takes the averages of the squared deviations. Finally, the rejection rate is defined as the share of replications where a  $t$ -test on the true parameter estimates rejects.

This small-scale Monte Carlo analysis allows to illustrate the source of the bias in the saturated employment-weighted OLS framework. Based on our discussion from above, we expect to obtain unbiased estimates for situations with no entry and exit (i.e.,  $p_n = 0$  and  $p_{i,x} = 0$ ) and for the experiments with a positive number of entrants but no exits (i.e.,  $p_n = 0.2$  and  $p_{i,x} = 0$ ). The reason for this is that in a saturated model entry is fully

---

<sup>7</sup>More discussion on the problem of regression to the mean effects is offered in Section 4.

captured by  $d_{i,1}$  and  $d_{i,2}$ , respectively. By contrast, for all situations with positive exit rates we expect the saturated and employment-weighted OLS estimator to be biased.

An additional issue to bear in mind is that the employment-weights themselves depend on  $g_{iT}$  and, thus, are endogenous as discussed in Footnote 2. The experiments without any entry and exit allows us to provide some first insights into this issue. If we observe any bias in this situation, it can only be triggered by the endogeneity of the weights  $w_{iT}$ . By contrast, under the data generating process imposed for this exercise the alternative maximum likelihood estimator should deliver unbiased estimates in all alternative experiments.

The first series of experiments assumes that  $\sigma = 0.1$ . In this case, extreme values of  $y_{iT}$  and  $y_{i0}$  exhibit low probability weights, respectively, and the linear approximation of  $g_{iT}$  discussed above should yield a small approximation error. Without entry and exit, i.e.,  $p_n = 0$  and  $p_{i,x} = 0$ , the employment-weighted estimator for the saturated model, in general, provides negligible biases and RMSEs and also delivers properly sized rejection rates which correspond to the nominal size of the t-test of 0.05. The parameter estimate and the rejection rate for the reference group captured by  $\beta_0$  forms a notable exception as documented in column (1) of Table 1. Accordingly, the marginal effect for the largest-oldest firms is overestimated also affecting the size of the resulting t-test. In line with the discussion above, this bias can only be triggered by the employment-weighting scheme that utilizes endogenous weights. This finding is reinforced by the results reported in column (5). Here, we increase  $\sigma$  from 0.1 to 0.5 resulting in an increase in the bias of the  $\beta_0$  estimate. More extreme values in  $y_{iT}$  and  $y_{i0}$  result in larger job creation rates  $g_{iT}$  which directly affect a firm's weight in the estimation procedure. Therefore, it seems to be that the endogenous weighting scheme only affects the estimates for the reference group. Indeed, when applying a simple non-weighted OLS estimator to the experiments reported in columns (1) and (5) one obtains unbiased estimates throughout. This reinforces that our finding is driven by the endogenous structure of the chosen weights in employment-weighted OLS regressions rather than by the approximation error.<sup>8</sup>

Turning our attention to the impact of firm entry and exit for the accuracy of the employment-weighted saturated OLS specification our main theoretical arguments are confirmed. All columns with positive exit rates (i.e.,  $p_{i,x} = 0.2$ ) indicate a (substantial) bias and increased RMSEs for all four parameters. An increase in  $\sigma$  severely affects the t-tests for the constant, while especially for  $\beta_2$  and  $\beta_3$  the simulated rejection rates are still

---

<sup>8</sup>This finding is also confirmed in additional experiments in which we apply random exogenous weights. The results from both additional simulation exercises are available from the authors upon request.



Table 1: Monte Carlo Simulations: 1,000 observations and 10,000 replications

	$\sigma=0.1$				$\sigma=0.5$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$p_n = 0$	$p_n = 0$	$p_n = 0.2$	$p_n = 0.2$	$p_n = 0$	$p_n = 0$	$p_n = 0.2$	$p_n = 0.2$
	$p_{i,x} = 0$	$p_{i,x} = 0.2$	$p_{i,x} = 0$	$p_{i,x} = 0.2$	$p_{i,x} = 0$	$p_{i,x} = 0.2$	$p_{i,x} = 0$	$p_{i,x} = 0.2$
<b>Bias*100</b>								
Employment-weighted OLS estimator using a saturated model for all observations and an entry dummy								
$\beta_0$	0.510	-19.837	0.508	-19.868	12.451	-8.100	12.421	-8.163
$\beta_1$	0.005	-3.610	0.007	-3.590	0.055	-3.434	0.081	-3.395
$\beta_2$	-0.033	-0.593	-0.033	-0.609	-0.184	-0.573	-0.192	-0.612
$\beta_3$	-0.023	0.470	-0.030	0.477	-0.115	0.402	-0.158	0.384
Maximum likelihood estimator for continuing firms only								
$\beta_0$	0.002	-0.003	0.001	-0.004	0.009	-0.015	0.005	-0.018
$\beta_1$	0.010	0.017	0.009	0.017	0.048	0.086	0.046	0.083
$\beta_2$	-0.013	-0.003	-0.015	-0.007	-0.063	-0.017	-0.075	-0.033
$\beta_3$	-0.005	-0.019	-0.005	-0.018	-0.025	-0.095	-0.023	-0.091
<b>Mean squared error*100</b>								
Employment-weighted OLS estimator using a saturated model for all observations and an entry dummy								
$\beta_0$	0.009	4.078	0.010	4.124	1.727	1.011	1.760	1.102
$\beta_1$	0.011	0.418	0.014	0.483	0.318	0.800	0.394	0.952
$\beta_2$	0.015	0.354	0.018	0.440	0.409	0.843	0.507	1.042
$\beta_3$	0.025	0.621	0.031	0.773	0.690	1.442	0.862	1.792
Maximum likelihood estimator for continuing firms only								
$\beta_0$	0.004	0.005	0.005	0.006	0.093	0.115	0.115	0.143
$\beta_1$	0.008	0.010	0.010	0.012	0.197	0.248	0.247	0.308
$\beta_2$	0.008	0.010	0.010	0.012	0.200	0.248	0.251	0.310
$\beta_3$	0.016	0.020	0.020	0.025	0.400	0.500	0.507	0.627
<b>Rejection rate</b>								
Employment-weighted OLS estimator using a saturated model for all observations and an entry dummy								
$\beta_0$	0.102	1.000	0.092	1.000	0.874	0.294	0.783	0.248
$\beta_1$	0.049	0.107	0.051	0.098	0.050	0.073	0.053	0.064
$\beta_2$	0.052	0.050	0.053	0.046	0.055	0.057	0.054	0.058
$\beta_3$	0.055	0.046	0.054	0.046	0.056	0.049	0.056	0.052
Maximum likelihood estimator for continuing firms only								
$\beta_0$	0.049	0.051	0.050	0.051	0.049	0.051	0.050	0.051
$\beta_1$	0.049	0.048	0.049	0.049	0.049	0.048	0.049	0.049
$\beta_2$	0.048	0.049	0.049	0.050	0.048	0.049	0.049	0.050
$\beta_3$	0.049	0.052	0.049	0.049	0.049	0.052	0.049	0.049

almost valid. By contrast, the results for entry without exit (i.e,  $p_n = 0.2$  and  $p_{i,x} = 0$ ) are virtually identical to the experiment with no entry and no exit. This confirms the view that a model which accounts for entry in a fully saturated way is able to substantially reduce the bias of the employment-weighted OLS estimator. Accordingly, the inclusion of an additional exit dummy together with a full set of its interactions with all firm size and age classes results in a linear three-part model and would allow to obtain consistent estimates when assuming exogenous employment weights.

Finally, the alternative maximum likelihood estimator (which is equivalent to an un-weighted OLS estimator for the log difference of continuing firms only) provides unbiased estimates and properly sized t-tests for all different experiments carried out. The RMSEs tend to increase with  $\sigma$ . This is well in line with standard results from the econometric literature. All these findings point to the attractiveness our the alternative three-part model for estimating DHS job creation rates.<sup>9</sup>

## 4 Who creates jobs in Austria?

We use data from the Austrian Social Security Database (ASSD) for our empirical application.<sup>10</sup> The ASSD is an administrative data set that includes a daily calender of employment relationships between individuals and firms and, thus, allows to calculate the (overall) number of employees in a respective firm at each point in time.<sup>11</sup> For our purposes, we calculate firm-specific annual employment figures in the non-farm business sectors over the time period from 1993 to 2009, taking June 7th as the reference day.

The resulting database comprises approximately 3 million firm-year observations, for which we apply the employment-weighted OLS estimator and the alternative three-part and multi-part models. We construct eight dummy variables for firm size and firm age, respectively and include a full set of interactions thereof. This results in 63 parameters to be estimated. We further include a full set of industry and year dummy variables. In

---

<sup>9</sup>The multi-part approach is not included in the Monte-Carlo analysis because, here we assume constant marginal effects across expanding and shrinking firms. However, in our empirical application we also compare the three-part model with the more general multi-part approach and are not able to identify remarkable differences across these alternative models.

<sup>10</sup>These data have been used extensively for empirical research in labor economics (see, e.g., Card *et al* 2007; del Bono *et al.* 2012) and industrial organization (see, e.g., Huber and Pfaffermayr 2010; Huber *et al* 2013).

<sup>11</sup>Fink *et al.* 2010 provide a comprehensive discussion on how to extract firm level information from the ASSD.

order to avoid regression to the mean effects we classify firms according to their average firm size in the years  $t$  and  $t - 1$ . This approach is in line with the previous literature on how to tabulate economic variables over consecutive time periods (see van de Stadt and Wansbeek 1990 for a formal treatment).

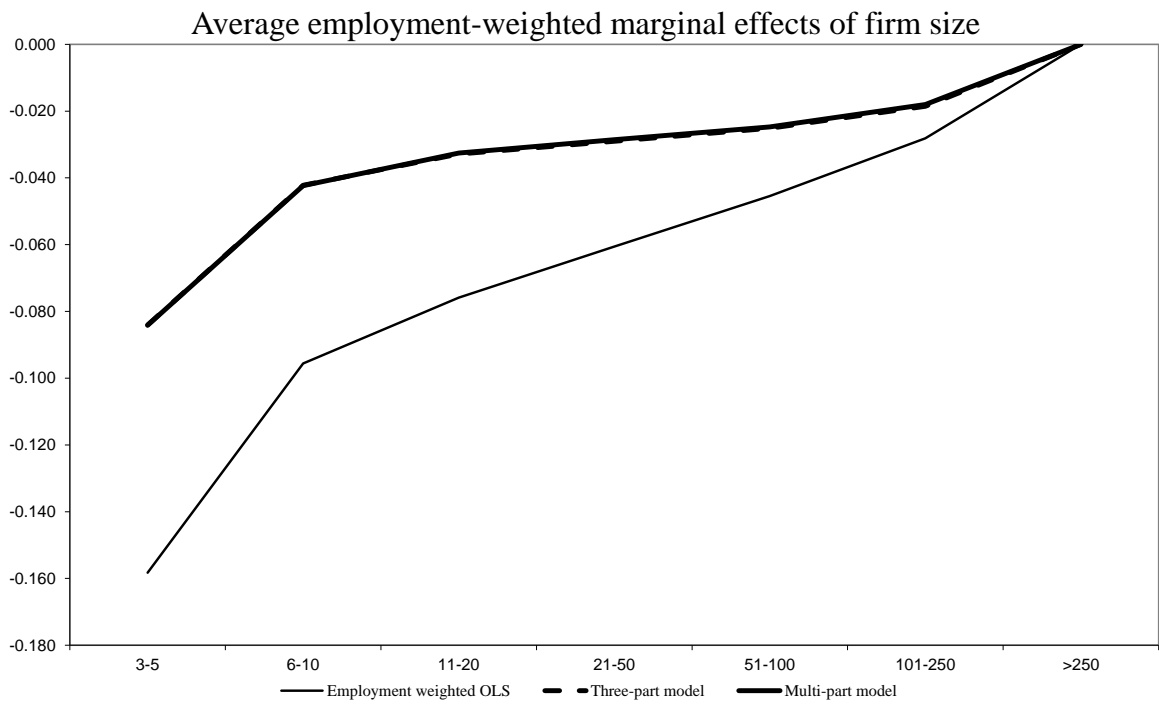
The empirical findings of this exercise are collected in Figure 1, which is based on the saturated model estimates discussed above.<sup>12</sup> The upper (lower) panel plots the average estimated (and employment-weighted) marginal effects of firm size (age) on the job creation rate. The marginal effects of firm size are calculated as the difference of predicted job creation rates and counterfactual ones where all firm size related dummy variables (including all interaction terms) are set to zero. Thus, age-, industry- and time effects are held constant when calculating the marginal firm size effects. The marginal firm age effects are computed analogously. This approach is similar to Haltiwanger *et al* (2013a) who evaluate the firm size (age) effects at the average within-cell firm age (size). The largest firms with more than 250 employees and firms with more than 20 years of age form the reference groups, respectively, and their marginal effects are normalized to one.

In general, both panels of Figure 1 clearly indicate that the employment weighted OLS estimates deviate substantially from their three- and multi-part counterparts. By contrast, the three-part and multi-part models deliver virtually identical marginal effects. The difference between the employment-weighted OLS estimator and its alternatives is most pronounced for the groups of smallest and youngest firms. The employment-weighted marginal effect for the smallest firms with less than 3 employees amounts to approximately -0.16 while it is only half that magnitude when estimated via the alternative maximum likelihood based estimators. In a similar vein, the employment-weighted OLS estimator delivers a slightly marginal effect for firms with one year of age (i.e., -0.003) while we alternatively estimate a marginal effect of around 0.083 to 0.087 when applying the maximum likelihood procedures.

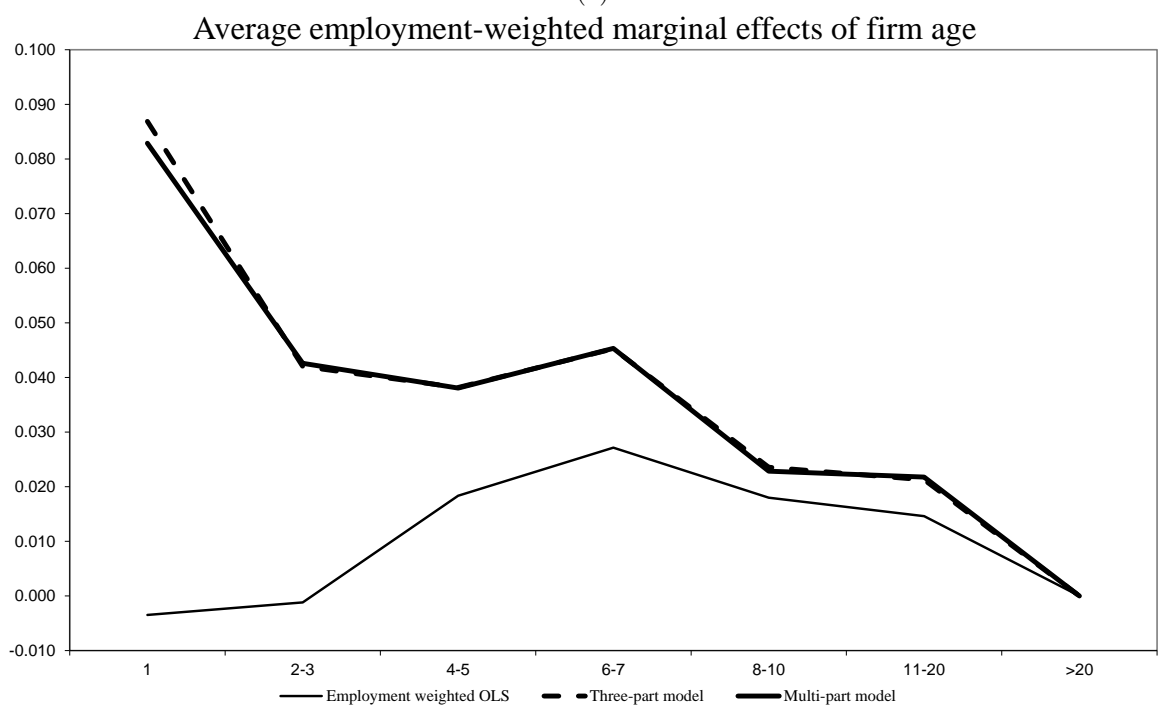
The reason for this is that exit hazards are largest for the smallest or youngest firms and, therefore, the pooling of entering, exiting and continuing firms most severely affects the job creation estimates within these two groups of firms. Moreover, equation (2) allows to discuss the observed differences in the estimates in econometric terms. Focussing on the lower panel of Figure 1, the bias from employment-weighted OLS is negative for all firm age classes. The reason for this is that by definition, firm entry is only possible with an age of zero years and, therefore, the entry probabilities are zero in all other firm age

---

<sup>12</sup>The actual parameter estimates are available upon request.



(a)



(b)

Figure 1: Average employment-weighted marginal effects of (a) firm size and (b) firm age.

classes. By contrast, firm exit is possible within all age cells. Accordingly, the second part of equation (2) is negative for all firm age cells and the resulting bias is unambiguously negative. In addition, the negative bias for the firm age effects decreases with firm age, which is consistent with the (descriptive) observation that exit probabilities decrease with firm age. This finding has also been highlighted in the literature on market exit.

To sum up, employment-weighted OLS estimation of DHS job creation rates unambiguously underestimates the marginal firm size and age effects on job creation rates. This underestimation is most pronounced for the youngest firms. For the Austrian economy employment-weighted OLS estimation also underestimates the contributions of the smallest firms to job creation. In general, this latter result depends on the entry and exit probabilities within the group of the smallest firms.

## 5 Conclusions

The analysis of (net) job creation by firms with different characteristics has a long tradition in economics. This discussion has been dominated by the question on which (small or large) firms are the most important net job creators. Only recently, Haltiwanger *et al.* (2013a) highlighted the role of firm age where their results indicate that young firms are the most crucial (net) creators of jobs.

From a methodological point of view, the incorporation of firm entry and firm exit in the analysis of job creation rates has been a serious challenge to empirical research. Davis and Haltiwanger (1992) and Davis, Haltiwanger and Schuh (1996) proposed a generalized measure of firm growth which permits such an integrated treatment. Their measure of job creation is defined for the closed interval of -2 to 2, where -2 (2) corresponds to firm entry (exit).

This paper analyzes the properties of econometric models for the DHS job creation rate. We specifically focus on one strand of the empirical literature which applies employment-weighted OLS estimation to a model that includes a full set of firm size and firm age dummies and their interaction terms. Since, this approach is equivalent to constructing cell-means as the unit of observations this paper's results are also applicable to the early job creation literature, which was limited to use age-size-industry-year cells as observational unit.

The main result of this paper is that employment-weighted OLS estimation of the DHS job creation rate delivers biased and inconsistent parameter estimates because even in saturated models one obtains a non-zero conditional expectation of the errors for exiting firms. The crucial mean independence assumption, thus, fails to hold due to the pooling of entering, exiting and continuing firms. The paper also derives an explicit formula which allows to quantify the bias. This shows that, the bias increases with the (absolute) net-entry rate and disappears if each estimated age-size cell exclusively contains either exiting, continuing or entering firms.

The paper also proposes alternative easy to implement three-part and multi-part estimation procedures that treat exiting, continuing and entering firms in separate models. Based on the assumption of a lognormal firm size distribution for the continuing firms we formulate a maximum-likelihood estimator that is equivalent to applying simple OLS estimation for employment growth measured in terms of the log difference for continuing firms only. This together with a probit model for exiting firms and accounting for job creation by entering firms constitute our three-part approach, which does not need to pool over all observations as in the case of the employment weighted regression for the DHS job creation rate with an entry dummy. The multi-part model combines multivariate probit estimation with two simple OLS estimators for the log-difference of shrinking and expanding firms, respectively.

A small-scale Monte Carlo analysis validates the theoretical findings of this paper: Employment-weighted OLS regressions deliver biased parameter estimates while the alternative three-part model is able to accurately deal with the DHS job creation rate as dependent variable. Finally, we apply the different estimators to the population of Austrian non-farm firms observed from 1993 to 2009 and compare the resulting estimates. For Austrian firms, the bias from employment-weighted OLS estimation is most pronounced for the groups of smallest and youngest firms. With regard to firm age, the empirical findings confirm the theoretical view that all employment-weighted job creation rate estimates are downward biased.

## References

- Angrist, Joshua D. and Jörn-Steffen Pischke (2009), *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, Princeton
- Armington, Catherine and Zoltan Acs (2004), Job Creation and Persistence in Services and Manufacturing, *Journal of Evolutionary Economics*, 14(3), pp. 309-325.

- Baldwin, John, Timothy Dunne and John Haltiwanger (1998), A Comparison of Job Creation and Job Destruction in Canada and the United States, *Review of Economics and Statistics*, 80(3), pp. 347-356.
- Birch, David L. (1979), *The Job Generation Process*, MIT Program on Neighborhood and Regional Change, Cambridge, MA.
- Bottazzi, Giulio and Angelo Secchi (2011), A New Class of Asymmetric Exponential Power Densities with Applications to Economics and Finance, *Industrial and Corporate Change* 20(4), pp. 991-1030.
- Burgess, Simon, Julia Lane and David Stevens (2000), The Reallocation of Labour and the Lifecycle of Firms, *Oxford Bulletin of Economics Statistics* 62(s1), pp. 885-907.
- Card, David, Raj Chetty and Andrea Weber (2007), Cash-on-Hand and Competing Models of Intertemporal Behavior: New Evidence from the Labor Market, *Quarterly Journal of Economics* 122(4), pp. 1511-1560.
- Capellari, Lorenzo and Stephen P. Jenkins (2003), Multivariate Probit Regression Using Simulated Maximum Likelihood, *Stata Journal* 3(3), pp. 278-294
- Caves, Richard E. (1998), Industrial Organization and New Findings on the Turnover and Mobility of Firms, *Journal of Economic Literature*, 36(4), pp. 1947-1982.
- Coad, Alex (2009), *The Growth of Firms: A Survey of Theories and Empirical Evidence*, Edward Elgar, Cheltenham.
- Davidson, Russell and James G. MacKinnon (1993), *Estimation and Inference in Econometrics*, Oxford University Press, Oxford, New York.
- Davis, Steven J. and John C. Haltiwanger (1992), Job Creation, Gross Job Destruction, and Employment Reallocation, *Quarterly Journal of Economics* 107(3), pp. 819-863.
- Davis, Steven J., John C. Haltiwanger and Scott Schuh (1996), *Job Creation and Destruction*, MIT Press, Cambridge MA.
- del Bono, Emilia, Andrea Weber and Rudolf Winter-Ebmer (2012), Clash of Career and Family: Fertility Decisions after Job Displacement, *Journal of the European Economic Association*, 10(4), pp. 659-683.
- Di Iorio, Francesca and Stefano Fachin (2004), Models of Labour Demand with Fixed Costs of Adjustment: A Generalised Tobit Approach, *Economics Bulletin*, 3(31), pp. 1-8.
- Dunne, Timothy, Mark J. Roberts and Larry Samuelson (1989), The Growth and Failure of U.S. Manufacturing Plants, *Quarterly Journal of Economics*, 104(4), pp. 671-698.
- Faberman, R. Jason (2003), *Job Flows and Establishment Characteristics: Variations Across U.S. Metropolitan Areas*, William Davidson Institute at the University of Michigan Stephen M. Ross Business School, William Davidson Institute Working Papers Series, No 610.

- Fink, Martina, Esther Kalkbrenner, Andrea Weber and Christine Zulehner (2010), Extracting Firm Information from Administrative Records: The ASSD Firm Panel, Working Paper 1004, NRN: The Austrian Center for Labor Economics and the Analysis of the Welfare State.
- Guertzgen, Nicole (2009), Firm Heterogeneity and Wages under Different Bargaining Regimes: Does a Centralised Union Care for Low-Productivity Firms?, *Jahrbuecher fuer Nationaloekonomie und Statistik*, 229(2-3), pp.239-253.
- Haltiwanger, John and Milan Vodopivec (2003), Worker Flows, Job Flows and Firm Wage Policies: An Analysis of Slovenia, *Economics of Transition* 11(2), pp. 253-290.
- Haltiwanger, John C., Ron S. Jarmin and Javier Miranda (2013a), Who Creates Jobs? Small vs. Large vs. Young, *Review of Economics and Statistics* 95(2), pp. 347-361.
- Haltiwanger, John C., Ron S. Jarmin and Javier Miranda (2013b), Web Appendix for Who Creates Jobs? Small vs. Large vs. Young, mimeo.
- Hart, Peter E. (2000), Theories of Firms' Growth and the Generation of Jobs, *Review of Industrial Organization* 17(3), pp. 229-248.
- Horrace, William C. and Ronald L. Oaxaca (2006), Results on the Bias and Inconsistency of Ordinary Least Squares for the Linear Probability Model, *Economics Letters*, 90(3), pp. 321-327.
- Huber, Peter and Michael Pfaffermayr (2010), Testing for Conditional Convergence in Variance and Skewness: The Firm Size Distribution Revisited, *Oxford Bulletin of Economics and Statistics* 72(5), pp. 648-668.
- Huber, Peter, Harald Oberhofer and Michael Pfaffermayr (2013), Job Creation and the Intra-distribution Dynamics of the Firm Size Distribution, *Industrial and Corporate Change*, forthcoming
- Ibsen, Rikke and Niels Westergaard-Nielsen (2005), Job Creation and Destruction over the Business Cycles and the Impact on Individual Job Flows in Denmark 1980-2001, *Allgemeines Statistisches Archiv/Journal of the German Statistical Society* 89(2), pp. 183-207.
- Ilmakunnas, Pekka and Mika Maliranta (2005), Worker Inflow, Outflow, and Churning, *Applied Economics* 37(10), pp. 1115-1133.
- Jovanovic, Boyan (1982), Selection and The Evolution of Industry, *Econometrica* 50(3), pp. 649-670.
- Komunjer, Ivana (2007), Asymmetric Power Distribution: Theory and Applications to Risk Measurement, *Journal of Applied Econometrics* 22(5), pp. 891-921.
- Lapatinas, Athanasios (2009), Labour Adjustment Costs: Estimation of a Dynamic Discrete Choice Model Using Panel Data for Greek Manufacturing Firms, *Labour Economics* 16(5), pp. 521-533.
- Moscarini, Guiseppe and Fabien Postel-Vinay (2012), The Contribution of Large and Small Employers to Job Creation in Times of High and Low Unemployment, *American Economic Review* 102(6), pp. 2509-2539.
- Neumark, David, Brandon Wall and Junfu Zhang (2011), Do Small Businesses Create More Jobs? New Evidence for the United States from the National Establishment Time Series, *Review of Economics and Statistics* 93(1), pp. 16-29.



- Nilsen, Øivind A., Kjell G. Salvanes and Fabio Schiantarelli (2017), Employment Changes, the Structure of Adjustment Costs, and Plant Size , *European Economic Review* 51(3), pp. 577-598.
- Searle, Shayle R. (1987), *Linear Models for Unbalanced Data*, Wiley, New York.
- Spletzer, James R. (2000), The Contribution of Establishment Births and Deaths to Employment Growth, *Journal of Business & Economic Statistics* 18(1), pp. 113-126.
- Stiglbauer, Alfred, Florian Stahl, Rudolf Winter-Ebmer and Josef Zweimüller (2003), Job Creation and Job Destruction in a Regulated Labor Market: The Case of Austria, *Empirica* 30(2), pp. 127-148.
- Sutton, John (1997), Gibrat's Legacy, *Journal of Economic Literature* 35(1), pp. 40-59.
- Tornqvist, Leo, Pentti Vartia and Yrjo O. Vartia (1985), How Should Relative Changes Be Measured?, *The American Statistician* 39(1), pp. 43-46.
- van de Stadt, Huib and Tom Wansbeek (1990), Miscellanea, Regression Effects in Tabulating From Panel Data, *Journal of Official Statistics* 6(3), pp. 311-317.
- Voulgaris Fotini, Theodore Papadogonas and George Agiomirgianakis (2005), Job Creation and Job Destruction in Greek Manufacturing, *Review of Development Economics* 9(2), pp. 289-301.
- Zhu Dongming and Victoria Zinde-Walsh (2009), Properties and Estimation of Asymmetric Exponential Power Distribution, *Journal of Econometrics* 148(1), pp. 86-99.

# Appendix

## The bias of the OLS estimator for the DHS job creation rate

In order to derive the bias of the weighted OLS estimator of the DHS job creation model that uses all observations, we partition the data into three groups of firms. Index  $x$  labels exiting firms (with  $y_{it} = 0$  and  $y_{i,t-1} \neq 0$ ) and index  $c$  refers to the continuing firms (with  $y_{it} \neq 0$  and  $y_{i,t-1} \neq 0$ ). The third set of observations with index  $n$  denotes entering firms (with  $y_{it} \neq 0$  and  $y_{i,t-1} = 0$ ).

For a weighted regressions one multiplies both sides of the estimating equation by some non-stochastic weight  $w_{it}$ . We define the matrix of weights  $\mathbf{W} = \text{diag}[\mathbf{W}_x, \mathbf{W}_c, \mathbf{W}_n]$ , where  $\mathbf{W}_k$ ,  $k = x, c, n$  comprise the weight of the respective observation in their main diagonal and have zero off-diagonal elements. Furthermore, we collect the indicators for exiting, continuing and entering firms in diagonal matrices  $\mathbf{D}_x, \mathbf{D}_c$  and  $\mathbf{D}_n$ , respectively. The resulting model is given then by

$$\begin{aligned} \mathbf{g} &= \mathbf{X}\beta + \mathbf{Z}_n\alpha + \mathbf{D}_n(2\mathbf{e} - \mathbf{X}\beta) + \mathbf{D}_x(-2\mathbf{e} - \mathbf{X}\beta) + \mathbf{D}_c\varepsilon \\ \mathbf{W}\mathbf{g} &= \mathbf{W}\mathbf{X}\beta + \mathbf{W}\mathbf{Z}_n\alpha + \mathbf{W}\mathbf{D}_n(2\mathbf{e} - \mathbf{X}\beta) + \mathbf{W}\mathbf{D}_x(-2\mathbf{e} - \mathbf{X}\beta) + \mathbf{W}\mathbf{D}_c\varepsilon \\ &= \mathbf{X}^*\beta + \mathbf{D}_n\mathbf{e}^*\alpha + \mathbf{D}_n(2\mathbf{e}^* - \mathbf{X}^*\beta - \mathbf{D}_n\mathbf{e}^*\alpha) + \mathbf{D}_x(-2\mathbf{e}^* - \mathbf{X}^*\beta) + \mathbf{D}_c\varepsilon^*. \end{aligned}$$

Here, the dummy design matrix for entering firms is denoted as  $\mathbf{Z}_n$  and  $\mathbf{e}$  is a vector of ones. Furthermore,  $\mathbf{Z}_n = \mathbf{D}_n\mathbf{e}$  and  $\mathbf{W}\mathbf{D}_k = \mathbf{D}_k\mathbf{W}$ ,  $k = x, c, n$  and the weighting scheme induces heteroskedastic disturbances so that estimation accounting for robust standard errors is called for.

For the model without an entry dummy, we obtain

$$\begin{aligned} \widehat{\beta} &= (\mathbf{X}^*\mathbf{X}^*)^{-1} \mathbf{X}^{*\prime} \mathbf{g}^* \\ &= \beta + (\mathbf{X}^*\mathbf{X}^*)^{-1} \mathbf{X}^{*\prime} (\mathbf{D}_n(2\mathbf{e}^* - \mathbf{X}^*\beta) + \mathbf{D}_x(-2\mathbf{e}^* - \mathbf{X}^*\beta) + \mathbf{D}_c\varepsilon^*) \end{aligned}$$

and

$$E[\widehat{\beta} - \beta | \mathbf{X}^*, \mathbf{D}_n, \mathbf{D}_c, \mathbf{D}_x] = -(\mathbf{X}^*\mathbf{X}^*)^{-1} \mathbf{X}^{*\prime} (2(\mathbf{D}_x - \mathbf{D}_n)\mathbf{e}^* + (\mathbf{D}_x + \mathbf{D}_n)\mathbf{X}^*\beta)$$

Using  $E[\mathbf{D}_c\boldsymbol{\varepsilon}|\mathbf{X}^*] = 0$  and applying the law of iterated expectations with  $E[\mathbf{D}_x|\mathbf{X}^*] = p_x\mathbf{I}$  and  $E[\mathbf{D}_n|\mathbf{X}^*] = p_n\mathbf{I}$  thus gives

$$E[\widehat{\beta} - \beta|\mathbf{X}^*] = -(p_x + p_n)\beta + 2(p_n - p_x)(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{e}^*$$

which is similar to the result derived for the saturated two-way model described in the text.

To analyze the model with an entry dummy let

$$\mathbf{M}_n^* = \mathbf{I}_n - \mathbf{Z}_n^*(\mathbf{Z}_n^{*\prime}\mathbf{Z}_n^*)^{-1}\mathbf{Z}_n^{*\prime}$$

and apply the Frisch, Waugh, Lovell theorem (see, e.g., Davidson and Mackinnon, 1993) to obtain

$$\begin{aligned}\widehat{\beta} &= (\mathbf{X}^{*\prime}\mathbf{M}_n^*\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{M}_n^*\mathbf{g}^* \\ &= \beta + (\mathbf{X}^{*\prime}\mathbf{M}_n^*\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{M}_n^*(\mathbf{D}_n(2\mathbf{e}^* - \mathbf{X}^*\beta - \mathbf{e}^*\alpha) + \mathbf{D}_x(-2\mathbf{e}^* - \mathbf{X}^*\beta) + \mathbf{D}_c\boldsymbol{\varepsilon}^*) \\ &= \beta + (\mathbf{X}^{*\prime}\mathbf{M}_n^*\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{M}_n^*((2 - \alpha)\mathbf{D}_n - 2\mathbf{D}_x)\mathbf{e}^* \\ &\quad - (\mathbf{X}^{*\prime}\mathbf{M}_n^*\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{M}_n^*(\mathbf{D}_n + \mathbf{D}_x)\mathbf{X}^*\beta + (\mathbf{X}^{*\prime}\mathbf{M}_n^*\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{M}_n^*\mathbf{D}_c\boldsymbol{\varepsilon}^*.\end{aligned}$$

Using  $\mathbf{M}_n^*\mathbf{Z}_n^* = \mathbf{M}_n^*\mathbf{D}_n\mathbf{e}^* = 0$ , assuming  $E[\mathbf{D}_c^*\boldsymbol{\varepsilon}|\mathbf{X}^*] = 0$ ,  $E[\mathbf{D}_k^*] = p_k\mathbf{I}_{NT}$ ,  $k = x, n$  and taking the expectation yields

$$\begin{aligned}E[\widehat{\beta} - \beta|\mathbf{X}^*, \mathbf{D}_n, \mathbf{D}_c, \mathbf{D}_x] &= -2(\mathbf{X}^{*\prime}\mathbf{M}_n^*\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{M}_n^*\mathbf{D}_x\mathbf{e}^* \\ &\quad - (\mathbf{X}^{*\prime}\mathbf{M}_n^*\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{M}_n^*(\mathbf{D}_n + \mathbf{D}_x)\mathbf{X}^*\beta.\end{aligned}$$

To analyze the consistency of the weighted OLS estimator for DHS growth rates we assume that the following limits exist for  $n = NT$  and  $T$  fixed:

1.  $|x_{ik}^*| < c$  for some constant  $c$ ,  $k = 1, \dots, K$ .
2.  $\lim_{n \rightarrow \infty} n^{-1}(\mathbf{X}^{*\prime}\mathbf{X}^*) = \mathbf{Q}$ , which is non-singular.
3.  $\lim_{n \rightarrow \infty} n^{-1}\mathbf{X}^{*\prime}\mathbf{e}^* = \mathbf{Q}_e$ .
4.  $p \lim_{n \rightarrow \infty} (n^{-1}\mathbf{X}^{*\prime}\mathbf{D}_c\boldsymbol{\varepsilon}^*) = 0$ .

Note that

$$\begin{aligned}\widehat{\beta} - \beta &= -2n(\mathbf{X}'\mathbf{X}^*)^{-1}(n^{-1}\mathbf{X}'(\mathbf{D}_x - \mathbf{D}_n)\mathbf{e}^*) + n(\mathbf{X}'\mathbf{X}^*)^{-1}n^{-1}\mathbf{X}'(\mathbf{D}_x + \mathbf{D}_n)\mathbf{X}^*\beta \\ &\quad + n(\mathbf{X}'\mathbf{X}^*)^{-1}(n^{-1}\mathbf{X}'\mathbf{D}_c\boldsymbol{\varepsilon}^*).\end{aligned}$$

Now  $d_{i,n}$  is assumed to be independent with  $P(d_{i,n} = 1) = p_n$ .  $\frac{1}{n}\mathbf{X}'\mathbf{D}_n\mathbf{e}^*$  has typical element  $\frac{1}{n}\sum_{i=1}^n x_{ik}^* e_i^* d_{i,n}$  with  $E[\frac{1}{n}\sum_{i=1}^n x_{ik}^* e_i^* d_{i,n}] = p_n \frac{1}{n}\sum_{i=1}^n x_{ik}^* e_i^*$  and  $Var[\frac{1}{n}\sum_{i=1}^n x_{ik}^* e_i^* d_{i,n}] = \frac{1}{n^2}\sum_{i=1}^n x_{ik}^{*2} e_i^{*2} p_n(1 - p_n)$ . Assuming  $|x_{ik}^*| < c$  for some constant  $c$  and using  $|e_i^*| < 1$ , it follows that  $\frac{1}{n^2}\sum_{i=1}^n x_{ik}^{*2} e_i^{*2} p_n(1 - p_n) \leq \frac{c^2}{n} p_n(1 - p_n)$ . Similar arguments hold for the typical element of  $n^{-1}\mathbf{X}'\mathbf{D}_x\mathbf{e}^*$  and  $\mathbf{X}'(\mathbf{D}_x + \mathbf{D}_n)\mathbf{X}^*\beta$ . Using Chebyshev's inequality and  $p\lim_{n \rightarrow \infty}(n^{-1}\mathbf{X}'\mathbf{D}_c\boldsymbol{\varepsilon}^*) = 0$  it follows that

$$p\lim_{n \rightarrow \infty} \left( \widehat{\beta} - \beta \right) = -2(p_x - p_n)\mathbf{Q}^{-1}\mathbf{Q}_e - (p_x + p_n)\beta$$

In the saturated model discussed in the text,  $\mathbf{X}^*$  is a dummy design matrix with  $H$  columns and one can verify that in this case  $n^{-1}(\mathbf{X}'\mathbf{X}^*) = n^{-1}(\mathbf{X}'\mathbf{W}^2\mathbf{X})$ ,  $n^{-1}\mathbf{X}'_x\mathbf{e}_x^* = n^{-1}(\mathbf{X}'\mathbf{W}^2\mathbf{e})$  and  $n(\mathbf{X}'\mathbf{W}^2\mathbf{X})^{-1}n^{-1}(\mathbf{X}'\mathbf{W}^2\mathbf{e}) = \mathbf{e}_H$ , where  $\mathbf{e}_H$  is a  $H \times 1$  vector of ones.

## Working Papers in Economics and Finance

### University of Salzburg

- 2013-05 **Peter Huber, Harald Oberhofer** and **Michael Pfaffermayr**. Who Creates Jobs? Estimating Job Creation Rates at the Firm Level
- 2013-04 **Jürgen Janger** and **Klaus Nowotny**. Career Choices in Academia
- 2013-03 **Klaus Nowotny**. Institutions and the Location Decisions of Highly Skilled Migrants to Europe
- 2013-02 **Sebastian Rathner**. The Industry-Specific Relationships between Corporate Financial Performance and 11 Corporate Social Performance Dimensions: Taking a More Nuanced Perspective
- 2013-01 **Sebastian Rathner**. The Relative Performance of Socially Responsible Investment Funds. New Evidence from Austria.
- 
- 2012-09 **Jörg Paetzold** and **Olaf van Vliet**. Convergence without hard criteria: Does EU soft law affect domestic unemployment protection schemes?
- 2012-08 **Martin Gächter, Peter Schwazer, Engelbert Theurl** and **Hannes Winner**. Regional density of private dentists: Empirical evidence from Austria.
- 2012-07 **Klaus Nowotny** and **Dieter Pennerstorfer**. Ethnic Networks and the Location Choice of Migrants in Europe.
- 2012-06 **Benjamin Furlan, Martin Gächter, Bob Krebs** and **Harald Oberhofer**. Democratization and Real Exchange Rates.
- 2012-05 **Peter Huber, Harald Oberhofer** and **Michael Pfaffermayr**. Job Creation and the Intra-distribution Dynamics of the Firm Size Distribution. Forthcoming in *Industrial and Corporate Change*.
- 2012-04 **Jörg Paetzold**. The Convergence of Welfare State Indicators in Europe: Evidence from Panel Data. Published in *European Journal of Social Security*.
- 2012-03 **Sebastian Rathner**. The Performance of Socially Responsible Investment Funds: A Meta-Analysis. Forthcoming as The Influence of Primary Study Characteristics on the Performance Differential Between Socially Responsible and Conventional Investment Funds: A Meta-Analysis in *Journal of Business Ethics*.
- 2012-02 **Jesus Crespo Cuaresma** and **Matthias Stöckl**. The Effect of Marketing Spending on Sales in the Premium Car Segment: New Evidence from Germany.

2012-01 **Harald Oberhofer, Matthias Stöckl and Hannes Winner.** The Wage Premium of Globalization: Evidence from European Mergers and Acquisitions. Published as The Wage Premium of Foreign Ownership: Evidence from European Mergers and Acquisitions in *Economics: The Open-Access, Open-Assessment E-Journal*.

---

2011-06 **Peter Huber.** The self-selection of Commuters.

2011-05 **Martin Gächter, Peter Schwazer, Engelbert Theurl and Hannes Winner.** Physician Density in a Two-Tiered Health Care System. Published in *Health Policy*.

2011-04 **Jesús Crespo Cuaresma and Max Roser.** Borders Redrawn: Measuring the Statistical Creation of International Trade. Published in *The World Economy*.

2011-03 **Harald Oberhofer and Michael Pfaffermayr.** FDI versus Exports: Multiple Host Countries and Empirical Evidence. Published in *The World Economy*.

2011-02 **Andrea M. Leiter, Magdalena Thöni and Hannes Winner.** Duo Cum Faciunt Idem, Non Est Idem: Evidence from Austrian Pain and Suffering Verdicts. Published as Pricing damages for pain and suffering in courts: The impact of the valuation method in *Journal of Empirical Legal Studies*.

2011-01 **Harald Oberhofer and Michael Pfaffermayr.** Testing the One-Part Fractional Response Model against an Alternative Two-Part Model.

---

2010-16 **Harald Oberhofer, Tassilo Philippovich and Hannes Winner.** Firm Survival in Professional Football: Evidence from the German Football League. Forthcoming in *Journal of Sports Economics*.

2010-15 **Engelbert Theurl and Hannes Winner.** The Male-Female Gap in Physician Earnings: Evidence from a Public Health Insurance System. Published in *Health Economics*.

2010-14 **Martin Feldkircher.** Forecast Combination and Bayesian Model Averaging - A Prior Sensitivity Analysis. Published in *Journal of Forecasting*.

2010-13 **Jesús Crespo Cuaresma and Octavio Fernández Amador.** Business Cycle Convergence in EMU: A Second Look at the Second Moment. Published in *Journal of International Money and Finance*.

2010-12 **Martin Feldkircher and Stefan Zeugner.** The Impact of Data Revisions on the Robustness of Growth Determinants - A Note on 'Determinants of Economic Growth: Will Data Tell?'. Published in *Journal of Applied Econometrics*.

2010-11 **Andrea M. Leiter, Magdalena Thöni and Hannes Winner.** Evaluating Human Life Using Court Decisions on Damages for Pain and Suffering. Published in *International Review of Law and Economics*.

- 2010-10 **Harald Oberhofer**. Employment Effects of Acquisitions: Evidence from Acquired European Firms. Published in *Review of Industrial Organization*.
- 2010-09 **Christian Reiner**. Regionale Arbeitsmärkte in der „Großen Rezession“: Dynamik regionaler Arbeitslosenquoten in Deutschland, Frankreich und Großbritannien im Krisenjahr 2009. Published in *Zeitschrift für Wirtschaftsgeographie*.
- 2010-08 **Leonardo Baccini** and **Andreas Dür**. The New Regionalism and Policy Interdependence. Published in *British Journal of Political Science*.
- 2010-07 **Harald Oberhofer** and **Michael Pfaffermayr**. Firm Growth in Multinational Corporate Groups. Published in *Empirical Economics*.
- 2010-06 **Sven P. Jost**, **Michael Pfaffermayr** and **Hannes Winner**. Transfer Pricing as a Tax Compliance Risk.
- 2010-05 **Christian Reiner**. Selling the Ivory Tower and Regional Development: Technology Transfer Offices as Mediators of University-Industry Linkages. Published as University policy and regional development: Technology transfer offices as facilitators and generators of university-industry linkages in *Berichte zur Deutschen Landeskunde*.
- 2010-04 **Matthias Stöckl**. Fremdkapitalquoten in Europa: Ein Ländervergleich. Published in *Wirtschaftspolitische Blätter*.
- 2010-03 **Jesús Crespo Cuaresma**, **Harald Oberhofer** and **Paul A. Raschky**. Oil and the Duration of Dictatorships. Published in *Public Choice*.
- 2010-02 **Matthias Stöckl** and **Hannes Winner**. Körperschaftsbesteuerung und Unternehmensverschuldung: Empirische Evidenz von europäischen Firmendaten. Published in *Journal of Economics and Statistics (Jahrbücher für Nationalökonomie und Statistik)*.
- 2010-01 **Andrea M. Leiter**, **Andrea M. Parolini** and **Hannes Winner**. Environmental Regulation and Investment: Evidence from European Country-Industry Data. Published in *Ecological Economics*
- 
- 2009-06 **Sven P. Jost**. Transfer Pricing Risk Awareness of Multinational Corporations: Evidence from a Global Survey.
- 2009-05 **Hannes Winner**. Der Kampf gegen internationale Steuerhinterziehung: Die OECD Initiativen gegen “Steueroasen”. Published in *Steuer und Wirtschaft*.
- 2009-04 **Michael Pfaffermayr**, **Matthias Stöckl** and **Hannes Winner**. Capital Structure, Corporate Taxation and Firm Age. Published in *Fiscal Studies*.
- 2009-03 **Simon Loretz** and **Padraig J. Moore**. Corporate Tax Competition Between Firms. Forthcoming in *International Tax and Public Finance*
- 2009-02 **Ronald W. McQuaid** and **Walter Scherrer**. Changing Reasons for Public Private Partnerships. Published in *Public Money and Management*.

2009-01 **Harald Oberhofer, Tassilo Philippovich and Hannes Winner.** Distance Matters in Away Games: Evidence from the German Football League. Published in *Journal of Economic Psychology*.