

RESEARCH SCHOOL OF INTERNATIONAL
TAXATION

TAX-RESPONSE HETEROGENEITY AND THE EFFECTS OF
DOUBLE TAXATION TREATIES ON THE
LOCATION CHOICES OF MULTINATIONAL FIRMS

SIMON BEHRENDT

GEORG WAMSER

WORKING PAPER
01/2018

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



SCHOOL OF BUSINESS AND
ECONOMICS

Tax-Response Heterogeneity and the Effects of Double Taxation Treaties on the Location Choices of Multinational Firms

Simon Behrendt¹ and Georg Wamser^{*2}

¹Zeppelin University

²University of Tuebingen

March 5, 2018

Abstract

This paper examines location choices of multinational enterprises (MNEs). We particularly focus on the consequences of double taxation treaties (DTTs) and corporate profit taxes on the probability to choose a location. DTTs have become a key policy instrument used by countries to regulate international tax issues related to the cross-border activities of MNEs. Based on three alternative location choice models, which all allow parameter estimates to vary randomly across firms, we show that firm responses to policy variables are highly heterogeneous. Postestimation statistics suggest that the heterogeneity of parameters is strongly correlated with firm size and effective tax burden, which is consistent with tax-avoidance behavior and provides an explanation for why tax-responses are heterogeneous in the first place. We quantify the (positive) effect of DTTs and demonstrate that the negative tax-responsiveness becomes larger if a DTT is implemented. The latter is evidence that provisions intended to prevent tax avoidance are effective.

Keywords: Location Choice, Multinational Firm, Double Taxation Treaties, Corporate Income Taxes

JEL: F23, H25, H26

*Corresponding author: University of Tuebingen, School of Business and Economics, CESifo and NoCeT, 72074 Tuebingen, Germany; georg.wamser@uni-tuebingen.de

1 Introduction

A large literature investigates the consequences of business taxes on the activities of firms. One of the main insights of this literature is that firms respond to taxes by adjusting at numerous margins of firm activity to avoid taxes. Some work has paid particular attention to the tax-responses of multinational enterprises (MNEs), whose activities span across many firm entities and countries (for a recent survey, see Egger and Stimmelmayer, 2017). A central research interest is to better understand the way how corporate taxes determine location choices of MNEs. Some recent contributions argue that the extent to which taxes determine MNEs' location choices depends on the extent to which firms are able to avoid taxes (e.g., Merlo *et al.*, 2016). If MNEs were fully able to avoid taxes – by reducing effective tax payments to zero – they would not be affected by the corporate tax in host countries, suggesting a zero tax-response (see, e.g., Egger *et al.*, 2014, analyzing investment in fixed assets). Hence, we expect that the different degrees to which future tax payments can be avoided produce a distribution of heterogeneous tax elasticities when firms decide where to locate their foreign entities.

This paper is concerned with the consequences of taxes and double taxation treaties (DTTs) on the location choices of MNEs. There are lots of anecdotes about policymakers who believe that issues of double taxation are a main barrier of foreign market entry. For example, in a recent visit of the Israeli prime minister to Australia, the two countries' leaders acknowledged that a prospective DTT will remove a major impediment Israeli companies face when doing business in Australia. In a joint statement, they declared that “they resolved to work towards concluding a Double Taxation Agreement which would remove tax impediments to bilateral economic activity and enhance the integrity of our respective tax systems”.¹ We add to previous studies in the following ways. First, we examine the determinants of location decisions of MNEs focusing on the *consequences* of *taxes*, where we allow estimated coefficients to vary randomly across MNEs, *DTTs*, and *interactions* of these two variables in discrete choice models. Especially the interactions are interesting as DTTs, on the one hand, facilitate foreign investments (a central objective of concluding a DTT is to ensure that income is taxed only once). On the other hand, DTTs often include information exchange agreements or provisions

¹<http://mfa.gov.il/MFA/PressRoom/2017/Pages/Joint-Statment-by-Prime-Ministers-of-Israel-and-Australia.aspx>; <http://www.aijac.org.au/news/article/partners-in-innovation>; accessed on Sept. 21, 2017.

related to international tax avoidance concerns. We argue that provisions – implemented to prevent tax avoidance and income shifting of MNEs – should generally lead to a higher effective tax burden and, for this reason, make firms more responsive to statutory corporate tax rates. Second, we estimate three alternative location choice models. Our empirical approach produces a distribution of estimated tax parameters where some firms strongly respond to taxes, while others do not respond at all. Third, based on this distribution, we demonstrate that postestimation statistics on firm characteristics can be used to learn about types of firms, given a firm’s estimated tax-responsiveness. Fourth, we shed light on the relationship between effective tax payments (from balance-sheet data) and the estimated tax-responsiveness. Fifth, we provide additional quantification results on the effects of DTTs.

To assess firms’ tax-responses we use *Orbis*, a micro-level dataset reporting the worldwide activities of MNEs. *Orbis* includes information on the location of foreign affiliates of MNEs as well as the location of global ultimate firm owners. The latter information is important as we are interested in the impact of bilateral DTTs implemented between host countries of foreign affiliates and the countries in which the global ultimate owners reside (as income is ultimately repatriated to the latter location). Based on our preferred specification, using a generalized multinomial logit model (G-MNL), we find (i) that tax-responses are highly heterogeneous and negative on average; (ii) that DTTs have a positive effect on location choice; interactions between DTTs and taxes are negative, which suggests that DTTs make it more difficult to shift profits and save taxes (e.g., through information exchange); (iii) that larger firms are less responsive to taxes. As for (ii), we provide a number of quantification results. For example, a DTT between Germany and the US increases the location choice probability of US MNEs to locate in Germany by about 6.5 percentage points, *ceteris paribus*. If the US and China conclude a DTT, the change in location probability is only about 1.7 percentage points. It seems that the long-run effects driven by the extensive margin of foreign activity, measured in terms of real effects such as employment, are substantial. Back-of-the-envelope calculations suggest that about 10 thousand jobs at US MNEs operating in the UK depend on whether the UK and the US agree on a DTT. Given that about half of our sample relates to greenfield investments, a realistic estimate of real job loss is probably about half this number (of course, without accounting for any efficiency effects). As for (iii), these findings are based on postestimation statistics, given the distribution of estimated tax parameters. It is consistent with earlier results suggesting that the size of an MNE is highly correlated

with the opportunities to shift profits and avoid taxes. We finally demonstrate that actual tax payments (relative to statutory tax rates) are smaller if a firm is rather unresponsive to tax incentives. In fact, low actual (effective) tax payments would explain why the firms are unresponsive in the first place, which confirms our basic argument that tax avoidance determines the tax parameter at the level of a firm.

Our paper is structured as follows. Section 2 briefly discusses the central mechanism producing heterogeneity in tax-responses. In Section 3 we present the two hypotheses we test in the empirical part on the consequences of DTTs. Section 4 introduces alternative discrete choice models. Section 5 formalizes a location choice model. Section 6 provides an overview on the different data we employ, including some description thereof. In Sections 7 to 9 we present the results of the empirical analysis. Especially Sections 8 and 9 look at the effects of DTTs on (heterogeneous) tax elasticities. We conclude in Section 10.

2 Heterogeneous tax-responses

This paper has two main objectives: (i) estimating and quantifying heterogeneous tax and DTT responses; and (ii) exploiting these estimates to learn about firms and firm behavior (and the assumptions producing heterogeneity in tax-responses in the first place). For the first objective – estimating heterogeneous tax-responses – we utilize random coefficient models to obtain a distribution of tax parameters, which reflect the heterogeneous responses to corporate taxes influencing the location decisions of MNEs.²

We argue that the responsiveness to a tax depends on the effective tax burden faced by firm entity i . To see this, let us denote the *effective average tax burden* of i by Θ_i , which is the relevant tax measure determining discrete choices of location (see, for example, Devereux and Griffith, 1998). The effective tax burden Θ_i depends on the statutory tax rate S_j , implemented in host country j , and the firm's tax base. The latter is affected through statutory rules at the level of countries, but is generally specific to firm i as it depends on the extent to which taxes can be avoided by i . Say the effective tax burden is expressed relative to before-tax

²In Section 7.1, we distinguish between more or less responsive firms along the distribution of tax parameters to learn about the types of firms and their behavior.

profits, which we denote by B_i . Then, we can formalize Θ_i as

$$\Theta_i = \frac{S_j \cdot (B_i - \phi_i B_i)}{B_i}. \quad (1)$$

The term $(B_i - \phi_i B_i)$ is the tax base. The share ϕ_i ($0 \leq \phi_i \leq 1$) captures the extent to which profits (tax base) can be reduced through tax avoidance, including income shifting, or even tax evasion.³ The fact that ϕ_i is usually not observed cannot be overcome by using actual tax payments or measures for effective tax burden from income statements. The reason is that taxes actually payed are the result of numerous endogenous firm responses to statutory rules, which ultimately determine $(B_i - \phi_i B_i)$, while for the analysis of location decisions we need variables that capture the incentive effects of the local tax systems.

To illustrate the effect of ϕ_i , let us assume that $\phi_i = 1$ (all taxes can be avoided). This suggests that the effective tax burden is zero, i.e. $\Theta_i = 0$. If income is fixed and $\phi_i = 0$, the tax is effectively a lump sum tax. Given a continuum of unobserved ϕ_i , for $i = 1, \dots, N$ foreign entities belong to an MNE, we hypothesize that the tax-responsiveness to statutory tax rules is heterogeneously distributed along with ϕ_i .

We may look at the implications of ϕ by using the definition of a tax semi-elasticity (and neglect indices for simplicity). Suppose the population response to taxes is given by the *generic* tax (semi-)elasticity $\tilde{\varepsilon}$. To be precise, we define $\tilde{\varepsilon} = \Delta P / (P \cdot \Delta \Theta)$, where P denotes the probability to choose a location and $\Delta \Theta$ denotes the change in the effective tax burden. Note that we expect $\tilde{\varepsilon}$ to be negative as an increase in Θ will make a location less attractive, as a larger share of profits is taxed. Hence, if a one percentage point increase in S leads only to a small increase in Θ , then the endogenous change in location probability $\Delta P / P$ will be small as well.⁴ Since the extent to which a change in S translates into a change in Θ depends on ϕ , this reasoning suggests heterogeneously distributed tax parameters in a population with heterogeneous tax avoidance, i.e. $\phi \in [0, 1]$.

³The parameter ϕ may include all forms of legal tax deductions, but the focus in our paper is on profit shifting and international tax planning. The latter is basically unobserved by tax authorities.

⁴Note that this argument applies to a fixed population elasticity $\tilde{\varepsilon}$.

3 The effects of DTTs on MNE activity

Over the last decades, many countries (or pairs of countries) have concluded DTTs. A DTT is an agreement between two jurisdictions to regulate cross-border tax issues. Blonigen and Davies (2004) suggest that DTTs may increase foreign direct investment (FDI) through a (i) clear definition of tax base to avoid double taxation; (ii) lower withholding taxes and less tax uncertainty related to unilateral changes in tax policy. Negative consequences of a DTT on FDI may be related to (iii) stricter transfer pricing regulation and information exchange between jurisdictions (which makes profit shifting activities of MNEs more difficult); (iv) specific provisions to prevent treaty shopping through operations in low-tax or tax haven affiliates. The role of DTTs can be summarized as follows: “From their inception the *raison d’être* of DTAs [double taxation agreements] has been the avoidance of double taxation. The solution to that problem necessarily involves taxing income only once and that leads to consideration of which country will have the taxing right. More recently, DTAs have also developed into instruments to prevent tax evasion in a cross-border context.” (Holmes, 2014, p. 58). This leads us to the two central hypotheses we analyze in this study. The first hypothesis suggests a positive impact of DTTs on foreign activity:

Hypothesis 1: *DTTs facilitate the cross-border operations of MNEs through double-taxation relief as well as clear definitions of relevant tax bases. Hence, we expect that a DTT between potential host country and ultimate owner country has a positive impact on the probability to choose a foreign location.*

We analyze Hypothesis 1 in a choice model about alternative locations, in which we allow the effect of DTTs to vary randomly across firms. Related to our hypothesis, but neglecting alternative choices, Davies *et al.* (2009), as well as Egger and Merlo (2011), confirm that DTTs induce a positive effect on the binary decision to invest or not.

Many studies, usually based on more aggregated data, suggest a negative effect of DTTs and ascribe the negative impact to additional provisions to prevent profit shifting (e.g., Blonigen and Davies, 2004; Egger *et al.*, 2006). We argue that stricter regulation of profit shifting and transfer pricing should mainly be reflected in the tax sensitivity of location choice. To see this, let us interpret our second hypothesis in light of Section 2 and the parameter ϕ_i introduced therein. Suppose a country-pair concludes a DTT, which includes provisions on information

exchange between tax agencies to prevent tax avoidance. This suggests that ϕ_i becomes smaller and the tax-responsiveness becomes larger (more negative), as implied by Equation (1) and its effect on the tax semi-elasticity explained above. Thus, we formulate our second hypothesis:

Hypothesis 2: *DTTs restrict tax avoidance activities of MNEs through information exchange or additional regulation. Hence, a DTT between potential host country and ultimate owner country makes location choices more responsive to statutory tax rates.*

Note that all DTTs usually include some form of provisions related to international tax avoidance concerns. For example, Luxembourg and Cyprus recently (May, 2017) signed a DTT. “This treaty incorporates the latest international standards as regards exchange of information and already fully takes into consideration the Base Erosion and Profit Shifting (“BEPS”) recommendations of the Organisation for Economic Co-operation and Development (“OECD”)” (see EY TAX Alert).⁵

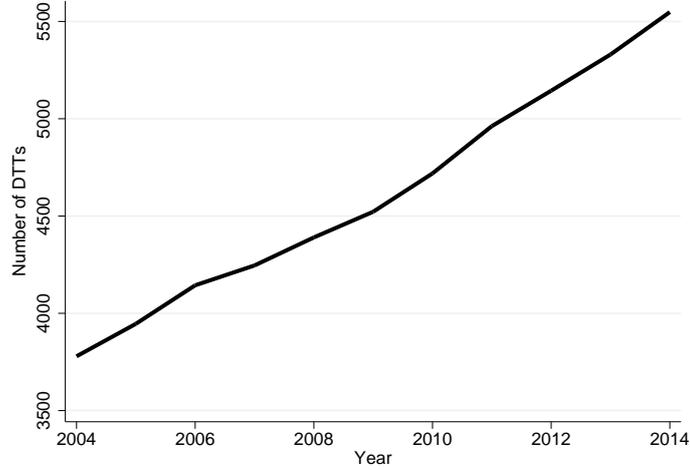
To address the two hypotheses, we use information from the IBFD’s tax research platform and define the variable DTT_{nj} , indicating whether country-pair nj has concluded a DTT ($DTT_{nj} = 1$) or not ($DTT_{nj} = 0$). Note that we use index n to denote MNE n making a choice about a potential host country j to set up a foreign affiliate later on.

In our data, of 34,938 possible country-pairs, 5,548 have concluded a DTT in the year 2014. Figure 1 suggests that the number of DTTs concluded between countries has increased by a factor of about 1.5 from 2004 to 2014. The cross-border investments of MNEs, and the additional tax-planning activities related to this, seem to be a major challenge for national tax systems. It appears that DTTs are one of the key policy instruments used to regulate cross-border activities of MNEs. The objective of this paper is therefore to provide new insights on the way how bilateral tax treaties affect the location choices of MNEs.

⁵Document downloadable under:
http://www.ey.com/Publication/vwLUAssets/DoubleTaxTreaty_between_Luxembourg_and_Cyprus/
(accessed on Sept. 21, 2017).

Figure 1: Number of DTTs over time

The figure shows the total number of DTTs concluded between pairs of countries from 2004 until 2014.



4 Alternative discrete choice models

When an MNE is choosing a location for one or several foreign affiliates, the firm essentially makes a discrete choice (Devereux and Maffini, 2007). We model such discrete choices by taking into account dynamics in location decisions over time and firm-specific unobserved heterogeneity. This section shortly introduces three models used in the subsequent empirical analysis: mixed logit (MIXL), scale heterogeneity logit (S-MNL), and generalized multinomial logit (G-MNL).

For the random coefficients derivation of MIXL, assuming a sample of N decision makers which make a choice from J alternatives in each of T choice situations, the utility that decision maker n obtains from choosing alternative j in choice situation (or period) t is given by

$$U_{njt} = \beta_n' x_{njt} + \varepsilon_{njt}, \quad (2)$$

where x_{njt} collects observed attributes of the decision maker and the alternative, β_n is a vector of coefficients of these attributes for decision maker n , and the random term ε_{njt} is assumed to be distributed i.i.d. extreme value. Utility can be rewritten as:

$$U_{njt} = (\beta + \eta_n)' x_{njt} + \varepsilon_{njt}. \quad (3)$$

Here, β represents the vector of mean utility weights, whereas η_n is the decision maker n -specific deviation from the mean. The coefficients vary over decision makers with density $f(\beta_n|\theta)$. This density is a function of parameters θ , which are estimated. Coefficients vary over decision makers but are assumed to be constant over choice situations for each decision maker. Let $y_{njt} = 1$ if decision maker n chooses alternative j in period t , and 0 otherwise. With β_n unknown, the unconditional choice probability of a sequence of choices $\{y_{njt}\}_{t=1}^T$ is the integral of the product of logit formulas over all values of β_n . In the case of a continuous mixing distribution, it is given by

$$P(\{y_{njt}\}_{t=1}^T) = \int \left[\prod_{t=1}^T \prod_{j=1}^J \left(\frac{e^{\beta_n' x_{njt}}}{\sum_k e^{\beta_n' x_{nkt}}} \right)^{y_{njt}} \right] f(\beta_n|\theta) d\beta_n. \quad (4)$$

Equation (4) has no closed form solution, but the parameters of $f(\beta_n|\theta)$ can be estimated by maximizing the simulated log-likelihood function:⁶

$$SLL = \sum_{n=1}^N \ln \left[\frac{1}{R} \sum_{r=1}^R \prod_{t=1}^T \prod_{j=1}^J \left(\frac{e^{\beta_{n[r]}' x_{njt}}}{\sum_k e^{\beta_{n[r]}' x_{nkt}}} \right)^{y_{njt}} \right] \quad (5)$$

where $\beta_{n[r]}$ denotes the r -th draw for decision maker n from $f(\beta_n|\theta)$.

However, Louviere *et al.* (1999, 2002) argue that scale heterogeneity is a major source of unobserved heterogeneity in discrete choice situations, i.e. a general scaling up or down of the entire coefficient vector. That is, with coefficients fixed, the scale of the random term in the multinomial logit (MNL) model is greater for some decision makers than for others. In MNL, the variance of the i.i.d. extreme value distributed random term ε_{njt} is given by $(\sigma^2\pi^2)/6$ and the scale parameter σ is usually normalized to one in order to achieve identification (Train, 2009). Following Fiebig *et al.* (2010), the MNL model can be rewritten with the scaling of the random term made explicit:

$$U_{njt} = \beta' x_{njt} + \frac{\varepsilon_{njt}}{\sigma} \quad (6)$$

Assuming that the scale parameter is heterogeneous in the population, its value for decision

⁶See Train (2009) for an extensive discussion of simulated maximum likelihood. Halton-draws are used in the maximization process of the simulated log-likelihood function.

maker n is denoted by σ_n :

$$U_{njt} = \boldsymbol{\beta}' \mathbf{x}_{njt} + \frac{\varepsilon_{njt}}{\sigma_n} \quad (7)$$

Here, the coefficients in the vector $\boldsymbol{\beta}$ are homogeneous across decision makers and all heterogeneity is in the variance of the random component of utility. S-MNL is obtained by multiplying with σ_n :

$$U_{njt} = (\sigma_n \boldsymbol{\beta})' \mathbf{x}_{njt} + \varepsilon_{njt} \quad (8)$$

The G-MNL model of Fiebig *et al.* (2010) allows to incorporate both a random coefficient vector and scale heterogeneity. Recent research finds that G-MNL is better at capturing choice behavior that is either lexicographic, i.e. choice is based mainly on a single observed attribute of an alternative, or random in the sense that it is only slightly influenced by observed attributes of an alternative (Keane and Wasi, 2013). G-MNL nests both MIXL (with normal mixing distribution) and S-MNL in two possible ways. G-MNL-I combines equations (3) and (8):

$$U_{njt} = (\sigma_n \boldsymbol{\beta} + \boldsymbol{\eta}_n)' \mathbf{x}_{njt} + \varepsilon_{njt} \quad (9)$$

G-MNL-II multiplies MIXL with σ_n :

$$U_{njt} = \sigma_n (\boldsymbol{\beta} + \boldsymbol{\eta}_n)' \mathbf{x}_{njt} + \varepsilon_{njt} \quad (10)$$

The G-MNL model is then given by

$$U_{njt} = (\sigma_n \boldsymbol{\beta} + \gamma \boldsymbol{\eta}_n + (1 - \gamma) \sigma_n \boldsymbol{\eta}_n)' \mathbf{x}_{njt} + \varepsilon_{njt}, \quad (11)$$

where the parameter γ determines how deviations from the means of the random coefficients are scaled. G-MNL-I results as a special case for $\gamma = 1$ and G-MNL-II for $\gamma = 0$. The distribution of σ_n is specified to be log-normal in order to constrain the scale parameter to be positive, so that $\ln(\sigma_n) \sim N(1, \tau^2)$, with $\tau > 0$ indicating scale heterogeneity. The random scale factor σ_n is specified as $\sigma_n = \exp(\bar{\sigma} + \tau \varepsilon_{0n})$, where $\varepsilon_{0n} \sim N(0, 1)$. For the model to be identified, $\boldsymbol{\beta}$ and τ are estimated and σ_n is calibrated to have an expected value of one. This is achieved by setting $\bar{\sigma} = -\tau^2/2$, as the expected value of the log-normally distributed σ_n is given by $E(\sigma_n) = \exp(\bar{\sigma} + \tau^2/2)$.

5 Location choice model

The effects of statutory corporate tax rates and DTTs on MNEs' location choices for foreign affiliates are modeled in the framework of a discrete location choice model. In this model, each possible location choice is associated with a latent payoff. In the following, an MNE, i.e. the decision maker in the location choice model, is defined as a firm that has at least one foreign affiliate. MNE n can choose one out of J potential host countries (locations) in period t to set up a foreign affiliate. In this case, each location $j = 1, \dots, J$ can be associated with the latent profit π_{njt}^* and MNE n bases its actual location choice in period t , $C_{njt} \in \{1, \dots, J\}$, on the maximum profit it can attain, $\text{argmax}\{\pi_{n1t}^*, \dots, \pi_{nJt}^*\}$. The location choice model states that latent profits in period t are a function of observable and unobservable country and firm characteristics

$$\pi_{njt}^* = (\alpha + \eta_n)TAXD_{njt} + \lambda DTT_{njt} + \delta SUB_{njt-1} + \beta' \mathbf{x}_{njt} + \varepsilon_{njt}, \quad (12)$$

where $TAXD_{njt}$ is the statutory corporate tax rate differential between potential host country j and the ultimate owner country in period t . We consider the tax rate differential, since not only the statutory corporate tax rate in the potential host country matters for the location choice, but rather how it compares to the tax rate in the ultimate owner country. DTT_{njt} indicates whether or not there exists a DTT between potential host country j and the ultimate owner country in period t , SUB_{njt-1} gives the number of affiliates the MNE already has in country j in period $t - 1$, \mathbf{x}_{njt} is a $(K \times 1)$ vector of country-level characteristics, and ε_{njt} is a disturbance term that captures unobserved firm and country effects. To account for unobserved heterogeneity in firms' response to taxes, the coefficient of the tax differential, $\alpha_n = \alpha + \eta_n$, is modeled as a firm-specific random coefficient. Assuming that the random coefficient follows a normal distribution with mean α and standard deviation σ to be estimated, i.e. $\alpha_n \sim N(\alpha, \sigma^2)$ or, equivalently, $\eta_n \sim N(0, \sigma^2)$, and ε_{njt} being distributed i.i.d. extreme value, yields the random parameters specification of MIXL, as described in the previous section. The parameters λ , δ , and β are fixed population parameters to be estimated.

While MIXL is the benchmark model in the following empirical analysis, we estimate two additional specifications of the location choice model. Scaling all coefficients in the location

choice model by σ_n introduces scale heterogeneity. This S-MNL specification is given by:

$$\pi_{njt}^* = \sigma_n(\alpha TAXD_{njt} + \lambda DTT_{njt} + \delta SUB_{njt-1} + \beta' \mathbf{x}_{njt}) + \varepsilon_{njt} \quad (13)$$

In addition, a G-MNL specification not only allows us to account for a random tax differential coefficient, but also for scale heterogeneity:

$$\begin{aligned} \pi_{njt}^* &= (\sigma_n \alpha + \gamma \eta_n + (1 - \gamma) \sigma_n \eta_n) TAXD_{njt} \\ &+ \sigma_n (\lambda DTT_{njt} + \delta SUB_{njt-1} + \beta' \mathbf{x}_{njt}) + \varepsilon_{njt} \end{aligned} \quad (14)$$

Thus, in equation (14), the coefficient of the tax differential is modeled as a random coefficient in the form of $\alpha_n = \sigma_n \alpha + \gamma \eta_n + (1 - \gamma) \sigma_n \eta_n$, with $\eta_n \sim N(0, \sigma^2)$. Since the deviations from the mean are equal to zero for λ , δ , and β , these coefficients are only scaled by σ_n . Apart from the specifications of the location choice model that involve a random coefficient for the tax differential, other specifications estimated in the following empirical analysis also allow random coefficients for DTT_{njt} and its interaction with $TAXD_{njt}$.⁷

6 Data

To empirically test the effects of statutory corporate tax rate differentials and DTTs on location decisions of MNEs, we use information from *Orbis*, a firm-level database provided by Bureau van Dijk.⁸ The empirical analysis covers a nine-year time period between 2004 and 2012. For each MNE in the sample, we record location choices over time. This is possible since the year and the location in which a new foreign affiliate is established are observed.⁹ It is important to note that we do not consider simultaneous location choices, i.e. more than one location is chosen in a year; if multiple foreign affiliates are established in one country and year, we take this as a single location choice. This leads to an unbalanced panel dataset with

⁷Note that none of the specifications include county fixed effects, since this would add a substantial number of parameters to be estimated. Maximization of the simulated log-likelihood function is numerically unstable, especially for G-MNL and S-MNL. Convergence depends on the number of parameters included in the model and both specifications failed to converge for such a large number of parameters.

⁸This dataset has been used before to study location choices of MNEs (for example, Arulampalam *et al.*, 2017). We have to note that it is not fully clear if new entries in *Orbis* are related to new location choices of MNEs, but we are quite confident that this is the case for the majority of observations as the median affiliate is established after the year 2001. Thus, most of the observations appear to be greenfield investment. We assume that all remaining new entries in our data are M&As and are counted as new location choices.

⁹MNEs in the dataset are identified as the sole owners of foreign affiliates or majority shareholders with more than 50% of shares.

MNEs making between one and eight location choices. The choice set in the location choice model consists of all J countries in which MNEs establish new foreign affiliates during the whole period from 2004 to 2012, excluding the respective country of residence of the MNE considered. The binary dependent variable y_{njt} is equal to one if firm n locates a foreign affiliate in country j in period t , and zero for the other $J - 1$ locations.

During the period under study a total of 4,961 MNEs are observed. These MNEs reside in one of 98 countries and establish their foreign affiliates in 83 locations, leading to an estimation sample of 1,007,383 observations with a total of 13,025 foreign affiliates being established.¹⁰ Out of these 4,961 MNEs, most establish two (2,545 MNEs) or three (1,254 MNEs) foreign affiliates between 2004 and 2012, while only a few establish just one (314 MNEs), seven (24 MNEs) or even eight (10 MNEs). Most foreign affiliates are established in Germany (2,937 entities) and the United Kingdom (1,687 entities). Other European countries like the Netherlands (758 entities), Poland (736 entities), Spain (546 entities), Austria (443 entities), France (388 entities), Luxembourg (374 entities), and Ireland (333 entities) are also important host countries. Nevertheless, emerging markets like Russia (840 entities), China (522 entities), and India (254 entities) attract a substantial number of foreign affiliates as well. Turning to ultimate owner countries, most MNEs are incorporated in the United States (819 entities), followed by the United Kingdom (337 entities), Luxembourg (337 entities), Germany (326 entities), the Netherlands (322 entities), and Switzerland (287 entities). An interesting fact about the dataset is that, in most cases, parent firms are the ones located in tax havens or financial offshore centers and not their foreign affiliates. This becomes important for the interpretation of some of our results.

Location choice is determined by a variety of factors. For each MNE the determinants of location choice are t -specific, where t is the year in which a new foreign affiliate is established. As stated in the previous section, $TAXD_{njt}$ measures the statutory corporate tax rate differential between country j and the country of residence of MNE n . The binary variable DTT_{njt} captures the effect of a DTT between foreign affiliate location j and the country of residence of MNE n . This variable is equal to one if there is a signed DTT between a country pair in period t , and zero otherwise. It is one key advantage of the Orbis dataset that we

¹⁰Due to missing values in some country-level independent variables for certain country-year combinations the estimation sample is smaller than the original sample. Moreover, since there exists no official economic data to construct the location-specific independent variables used in the following empirical analysis for some typical tax haven countries, these countries have to be discarded as possible locations for foreign affiliates.

observe many n - j -specific combinations, which is usually not the case in other datasets, to identify the effects of DTTs. Dynamics are introduced in the location choice model through the variable SUB_{njt-1} . This variable counts the number of foreign affiliates an MNE has in different locations in period $t - 1$. One should note that this is different from including a lag in the location choice of MNEs in the analysis. In case of SUB_{njt-1} , MNEs can have foreign affiliates in multiple countries in the previous period and the variable does not follow a binary pattern. Thus, endogeneity considerations that apply for lagged dependent variables in panel data context are not considered in the following.

Additional country-specific variables are included in the analysis. TCR_{jt} is a binary variable that is equal to one if country j imposes a strict thin capitalization rule, and zero otherwise. While the data on TCRs is taken from Merlo *et al.* (2016), we define TCR_{jt} in the following way. TCR_{jt} is equal to one only if a country is among the 10% of countries that have implemented the strictest TCRs, given the distribution of safe-haven ratios presented in Merlo *et al.* (2016). The reason for focussing only on the 10% strictest rules is that, while many countries have TCR law, the actual rules are rather lax in many countries and the jump to the next, less strict TCR is quite substantial. Ceteris paribus, we expect a negative sign for TCR_{jt} . As a proxy for market size and demand conditions in country j , the logarithm of a country's GDP, $\log(GDP)_{jt}$, is included. This variable is expected to have a positive effect. We further employ the logarithm of the GDP per capita in country j , $\log(GDPC)_{jt}$. A positive effect of this variable is expected if the foreign affiliate is part of a horizontal FDI strategy, as far as $\log(GDPC)_{jt}$ is positively related to purchasing power. By contrast, a higher GDP per capita might be linked to higher average wages, which would negatively affect the location choice probability if the foreign affiliate is part of a vertical FDI strategy and the MNE produces intermediate goods in low wage countries. GDP growth in country j is measured by $GDPGR_{jt}$. This variable captures the general economic attractiveness of a location and is, ceteris paribus, expected to have a positive effect. Since establishing an affiliate in a foreign country entails fixed investment costs, the variable $COSTBS_{jt}$, measuring costs of business start-up procedures relative to GNI per capita in country j , is included in the analysis. Ceteris paribus, higher investment costs are expected to decrease location choice probability. The effect of country j 's inflation rate on the probability of being chosen as a location is assessed by including the variable $INFL_{jt}$. Institutional quality in country j is measured by two variables: a freedom from corruption index ($CORRUPT_{jt}$) and a property

rights index ($PRIGHT_{jt}$). Both indices take on values between 0 and 100, with higher values indicating less corruption or higher standards of property rights, respectively. One would assume that less corruption and higher property rights are favorable characteristics of a potential host country. Thus, a positive sign is expected for both variables. Other variables account for geographical and cultural distance between potential host country and ultimate owner country. These variables do not change over the period under study for any country pair and thus the subscript t is omitted. The logarithm of the distance (in kilometers) between the most populated cities of potential host country and the ultimate owner country is captured by $\log(DIST)_{nj}$. The binary variable $CONTIG_{nj}$ is equal to one if both countries share a common border, and zero otherwise. A smaller geographical distance between potential host country and ultimate owner country is expected to have a positive effect. Cultural distance is captured by $COLONY_{nj}$, which is a binary variable that is equal to one in case country j and the country where the MNE is incorporated share a colonial past, and zero otherwise. Furthermore, $COMLANG_{nj}$ is a binary variable equal to one if both countries share a common (official) language, and zero otherwise. Table 1 contains definitions and summary statistics for all explanatory variables. For a comprehensive list of data sources, see Table A.1 in the appendix.

Table 1: Variable definitions and summary statistics

Summary statistics are based on the whole estimation sample, which contains 1,007,383 observations. Country-level characteristics are given for all potential host countries from the choice set and are measured in the year a new foreign affiliate is established. Binary measures are given for all country pairs of ultimate owner country (where MNE n is incorporated) and potential host countries $j = 1, 2, \dots, J$. GDP is measured in constant PPP US Dollars, using 2005 as base year.

Variable	Definition	Min	Max	Mean	Std. dev.
$TAXD_{njt}$	Statutory corporate tax rate differential between country j and ultimate owner country	-0.550	0.420	-0.030	0.130
DTT_{njt}	Binary variable indicating whether a country pair (ultimate owner country and host location) has signed a DTT	0.000	1.000	0.614	0.487
$TAXD-DTT_{njt}$	Interaction term of $TAXD$ and DTT	-0.450	0.400	-0.029	0.085
SUB_{njt-1}	Indicator of the number of affiliates MNE n has in country j in $t - 1$	0.000	228.000	0.047	0.684
TCR_{jt}	Binary variable indicating strict TCR rules in country j (within 10th percentile)	0.000	1.000	0.095	0.293
$\log(GDP)_{jt}$	(log of) Gross domestic product (GDP) in country j	22.273	30.235	25.860	1.701
$\log(GDPPC)_{jt}$	(log of) Gross domestic product per capita (GDP per capita) in country j	6.932	11.212	9.474	0.922
$GDPGR_{jt}$	Gross domestic product growth (GDP growth) in country j	-17.955	18.869	3.614	4.109
$COSTBS_{jt}$	Cost of business start-up procedures (% of GNI per capita) in country j	0.000	195.500	19.046	28.203
$INFL_{jt}$	Average consumer prices percentage change (inflation) in country j	-4.480	59.220	4.824	4.857
$CORRUPT_{jt}$	Freedom of corruption index of country j (scale ranges from 0-100; higher values indicating less corruption)	10.000	97.000	51.134	23.683
$PRIGHT_{jt}$	Property rights index of country j (scale ranges from 0-100; higher values indicating better property rights)	0.000	95.000	57.123	24.674
$\log(DIST)_{nj}$	(log of) Distance in kilometers between the most populated cities of a country pair	4.088	9.901	8.418	1.005
$CONTIG_{nj}$	Binary variable indicating whether a country pair shares a common border	0.000	1.000	0.033	0.179
$COLONY_{nj}$	Binary variable indicating whether a country pair ever had a colonial relationship	0.000	1.000	0.045	0.207
$COMLANG_{nj}$	Binary variable indicating whether a country pair shares a common (official) language	0.000	1.000	0.146	0.353

7 Basic results

Different specifications of the location choice model are estimated to determine the effects of statutory corporate tax rate differentials and bilateral DTTs on the location choice probability of potential host countries. In Table 2, the estimation of the heterogeneous response to tax rate differentials is of primary interest. Thus, in the MIXL and G-MNL specifications, the coefficient of $TAXD_{njt}$ is modeled as a random coefficient (following a normal distribution in the case of the MIXL specification).¹¹ In addition, this analysis serves as a horse race to see which of the three discrete choice models introduced above provides the best fit for the empirical estimation of the location choice model.

The estimated mean of $TAXD_{njt}$ is negative and statistically significant at the 1%-level across all model specifications: a higher tax at the host location relative to the home location makes a host country less attractive. Furthermore, the estimated standard deviations of $TAXD_{njt}$ in the MIXL and G-MNL specifications are highly significant, which suggests that there is indeed some heterogeneity in how tax rate differentials affect location choices of MNEs. In the context of MIXL with normally distributed $TAXD_{njt}$ coefficient, the estimated mean and standard deviation of the coefficient provide information on the share of the population of MNEs that places a negative value on this location attribute. Here, for roughly 64% of MNEs, the impact on their location choice is negative, while for the remaining 36% of MNEs the effect is positive.¹² Given that many MNEs are located in low-tax or tax haven countries, a positive estimate on $TAXD_{njt}$ may reflect a competitive advantage for these MNEs, compared to MNEs in high-tax countries. The former may simply avoid taxes to a large extent through the shifting of tax base to the ultimate owner location. The coefficients of DTT_{njt} , SUB_{njt-1} , and TCR_{jt} are also statistically significant at the 1%-level and have the expected signs. Thus, *ceteris paribus*, a bilateral DTT increases the location choice probability (we will look into this effect in more detail below); a larger number of previous investments in a country also increases the probability to choose a certain location; strict TCRs in a potential host country have a negative effect. The estimated coefficients of the other control variables are largely in line with what is expected and the results can be summarized as follows. First,

¹¹Note that the previous paper by Griffith *et al.* (2014) has used a random coefficient approach for the tax variable as well.

¹²These numbers are calculated the following way: $100 \cdot \Phi(-\hat{\beta}_j/\hat{s}_j)$, where Φ is the cumulative standard normal distribution and $\hat{\beta}_j$ and \hat{s}_j are the estimated mean and standard deviation of the j^{th} coefficient, respectively.

a larger market size, favorable demand conditions, and a greater general economic attractiveness, captured by $\log(GDP)_{jt}$ and $GDPGR_{jt}$, have a positive effect on the location choice probability. On the other hand, a higher GDP per capita ($\log(GDPC)_{jt}$) has a clear negative effect on the location choice probability. The same is true for higher fixed investment costs ($COSTBS_{jt}$). A country with a better protection of property rights ($PRIGHT_{jt}$) is more likely to be chosen as a location for foreign affiliates. The same is the case for countries that are closer in terms of geographical distance ($\log(DIST)_{nj}$) and historic ties ($COLONY_{nj}$).

Table 2: Basic estimation results

Dependent variable is the location choice indicator. Estimation is based on the whole sample of 1,007,383 observations. In MIXL and G-MNL, the coefficient of $TAXD_{njt}$ is defined as random, other coefficients are fixed. Simulation is based on 500 Halton-draws and robust standard errors are reported.

	(1)		(2)		(3)	
	MIXL		S-MNL		G-MNL	
	Estimate	Std. err.	Estimate	Std. err.	Estimate	Std. err.
$TAXD_{njt}$ (mean)	-1.319***	0.251	-2.401***	0.312	-1.785***	0.344
$TAXD_{njt}$ (std. dev.)	3.729***	0.320			2.023***	0.496
DTT_{njt}	1.072***	0.056	1.484***	0.090	1.497***	0.088
SUB_{njt-1}	0.934***	0.030	1.505***	0.038	1.510***	.039
TCR_{jt}	-0.165***	0.031	-0.276***	0.039	-0.286***	0.040
$\log(GDP)_{jt}$	0.544***	0.014	0.633***	0.020	0.639***	0.020
$\log(GDPC)_{jt}$	-0.110***	0.041	-0.099*	0.054	-0.123**	0.055
$GDPGR_{jt}$	0.019***	0.005	0.017**	0.007	0.019***	0.007
$COSTBS_{jt}$	-0.012***	0.001	-0.015***	0.002	-0.016***	0.002
$INFL_{jt}$	0.001	0.003	-0.004	0.005	-0.002	0.005
$CORRUPT_{jt}$	-0.001	0.001	-0.003	0.002	-0.002	0.002
$PRIGHT_{jt}$	0.015***	0.001	0.017***	0.002	0.016***	0.002
$\log(DIST)_{nj}$	-0.598***	0.015	-0.652***	0.017	-0.678***	0.018
$CONTIG_{nj}$	0.090**	0.043	0.046	0.048	0.030	0.049
$COLONY_{nj}$	0.466***	0.038	0.464***	0.041	0.486***	0.041
$COMLANG_{nj}$	0.061*	0.036	0.032	0.044	0.025	0.044
τ			0.665***	0.025	0.676***	0.025
γ					-3.019***	0.998
AIC	71,774.53		69,043.66		69,004.86	
BIC	71,963.69		69,232.83		69,217.67	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The positive and highly significant estimates of τ for S-MNL and G-MNL indicate the presence of scale heterogeneity, which MIXL fails to capture. Moreover, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) point to S-MNL and G-MNL as superior models in terms of model fit, with G-MNL providing the best model fit overall. However, the differences of S-MNL and G-MNL in terms of model fit are not large, compared to the improvement these models provide relative to the MIXL specification. The G-MNL

specification is not close to either G-MNL-I or G-MNL-II, according to the estimate of γ , which is smaller than zero and highly significant.

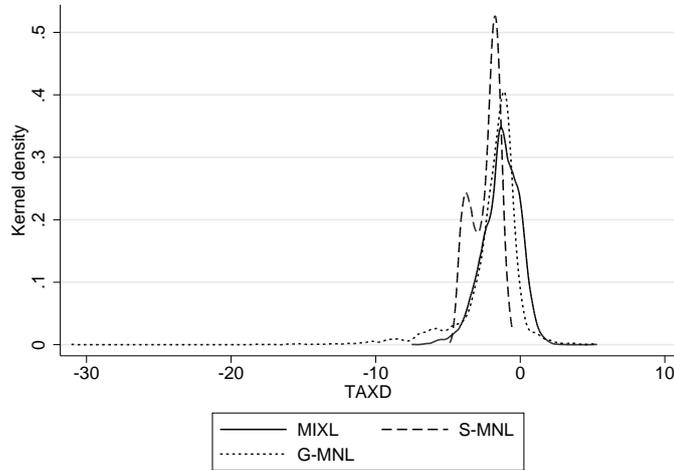
Knowing that S-MNL and G-MNL provide a better model fit than MIXL is not all that matters. It is also important to understand which aspects of firm behavior they capture better by examining the distribution of firm-specific heterogeneity with respect to taxes. Adopting an approach closely related to the “approximate Bayesian” approach of Allenby and Rossi (1998), individual-level $TAXD_n$ coefficients can be calculated using maximum likelihood estimation. Following Revelt and Train (2000), the central concept is a distinction between the distribution of preferences in the population of MNEs, and the distribution of preferences in the sub-population of MNEs that make particular location choices. In a first step, maximum likelihood estimation is used to estimate the distribution of preferences in the population using the data for all sampled MNEs. Then, the distribution of preferences of each sampled MNE is derived, conditional on the observed data for that MNE and the estimated population distribution of preferences. Intuitively, individual-level tax coefficients can be thought of as estimating the conditional mean of the coefficient distribution for the sub-group of MNEs facing the same location alternatives and making the same choices. Figure 2 plots the distributions of the individual-level (or firm-level) $TAXD_n$ coefficients.

We can see that the distribution of MIXL individual-level coefficients has a shape closely resembling a normal distribution. Allenby and Rossi (1998) note that MIXL with normally distributed coefficients has the tendency to draw in outliers, which is why the model has difficulties to capture “extreme” cases of MNEs placing great weight on tax rates. By contrast, the distributions of individual-level tax coefficients of S-MNL and G-MNL depart substantially from normality. They generate a mass of firms in the left tail that care intensely about tax differentials in location choices. These results shed more light on behavioral aspects of MNEs’ location choices. While higher differentials have a negative effect on the location choice probability for most MNEs, there seem to be MNEs that place an exceptionally great weight on corporate tax incentives and thus are more likely to choose locations with very low statutory corporate tax rates (i.e., low or negative tax rate differentials between potential host and ultimate owner countries). Some of these MNEs might engage in tax avoidance strategies that make use of tax differentials between the high-tax country, where the MNE is incorporated, and the low-tax countries, where the foreign affiliates are established. S-MNL and G-MNL

are able to capture these behavioral aspects in location choices of MNEs, which cannot be modeled within the MIXL framework. Nonetheless, the interpretation of individual-level tax coefficients is certainly limited in this paper, as the panel dataset is highly unbalanced with MNEs facing a fixed choice set of countries and with the time dimension T being relatively low for most firms, which negatively affects the properties of the estimated conditional means (for a detailed technical discussion, see Revelt and Train, 2000; Train, 2009).

Figure 2: Individual-level $TAXD_n$ coefficients

The figure plots Kernel densities of the individual-level $TAXD_n$ coefficients obtained from the MIXL, S-MNL and G-MNL specifications in Table 2 (here indicated by $TAXD$). Simulation is based on 500 Halton-draws.



Based on the preferred G-MNL specification in Table 2, we further investigate the effect of DTTs on the location choice probability. Since most MNEs in our sample are incorporated in the US, we focus on US firms and four potential host countries: Germany and the United Kingdom (the locations chosen most often in our dataset), as well as Russia and China (Non-European locations chosen most often in our dataset). Table 3 depicts the average location choice probabilities for US firms when considering these four host countries and the two cases where DTT_{njt} is either set equal to 0 or 1 for all t . As expected from the results of Table 2, a DTT leads to an increase in the average location choice probability for all four potential host countries, where the increase is largest for the United Kingdom (8.718%) and smallest for Russia (1.019%).

Table 3: Effect of DTTs on the location choice probability for US firms

The table reports average location choice probabilities (in percent) for the potential host countries China, Germany, the United Kingdom, and Russia. The ultimate owner country corresponds to the US. Two scenarios are illustrated: an existing DTT ($DTT = 1$) between potential host country and ultimate owner country as well as the absence of such a bilateral DTT ($DTT = 0$). Calculations are based on specification (3) in Table 2 and 500 Halton-draws.

	<i>Potential host country</i>			
	Germany	United Kingdom	Russia	China
$DTT = 0$	7.351	12.627	0.862	0.885
$DTT = 1$	13.908	21.345	1.881	2.567

We can use the estimates presented in Table 3 as well as the *Orbis* data to learn about the implications of a DTT in terms of number of employees (given the location choice sensitivity of firms). Of course, what we can provide here are only rough approximations of long-run employment effects using the predictions from our model. For the calculation, we first take the number of US-held affiliates that newly enter our dataset between 2006 and 2012 (denote this number by NA). We require the new affiliates to stay in the dataset until 2012, i.e. those entities that exit the market within this time span are not taken into account. NA is 860 (Germany), 1,154 (Great Britain), 179 (Russia), and 199 (China). We then multiply this number by the median number of employees of these new affiliates (ME) and by the change in base probability associated with concluding a DTT ($\Delta\widehat{BP}$) from Table 3. Table 4 provides the results of our quantification, including the total effect of a DTT on the number of employees (\widehat{E}). The latter is calculated as $\widehat{E} = NA \cdot ME \cdot \Delta\widehat{BP}$ (see last column in Table 4). Note that the estimates should be interpreted as long-run effects, which materialize over one or two decades. Anyway, they seem to be quite substantial: thousands of jobs at US MNEs hinge on whether a bilateral DTT is in place or not.

Table 4: Quantification of employment effects in host countries

The table reports employment effects, given the estimated effect of a DTT in Table 3, for the potential host countries China, Germany, the United Kingdom, and Russia. The ultimate owner country corresponds to the US.

	NA	ME	$\Delta\widehat{BP}$	\widehat{E}
Germany	860	112.0	0.06557	6,316
United Kingdom	1,154	101.0	0.08718	10,161
Russia	179	141.5	0.01019	258
China	199	270.0	0.01682	904

We additionally estimate MIXL and G-MNL specifications with random DTT_{njt} coefficient, again following a normal distribution in the case of MIXL. The justification for specifying DTT_{njt} as random may be based on the same arguments as in the case of $TAXD_{njt}$. Results are presented in Table 5. As before, the mean coefficient of DTT_{njt} is positive and statistically significant at the 1%-level. Moreover, there seems to be unobserved heterogeneity across firms not only with respect to taxes, but also regarding DTTs, since the estimated standard deviation of the DTT_{njt} coefficient is also highly statistically significant.¹³ Based on the MIXL specification, only 3% of MNEs prefer no bilateral DTT between host and ultimate owner country when choosing a location, which clearly confirms Hypothesis 1 that DTTs have a positive impact on the probability to choose a foreign location for most firms.

Again, based on AIC and BIC, the G-MNL specification provides the best model fit and there is empirical evidence for scale heterogeneity. However, the statistically highly significant estimate of γ is now markedly smaller in size than in the basic G-MNL specification presented in Table 2. Still, the estimated model is not close to either the G-MNL-I or G-MNL-II representation. Overall, we find evidence for tax-response heterogeneity of MNEs, which is consistent across all model specifications. The effect of DTTs on MNEs' location choices also seems to be heterogeneous but mostly positive. The models incorporating scale heterogeneity perform better than a MIXL model, which only allows for random coefficients.

¹³We also estimated a specification where the coefficient of DTT_{njt} follows a log-normal distribution to impose a sign restriction, i.e. restricting the effect of DTTs to be positive. However, the estimated standard deviation of the coefficient is not statistically significant in this specification, while it is when not imposing the sign restriction. To allow for both positive and negative effects of DTTs on location choice probabilities, we assume the coefficient of DTT_{njt} to follow a normal distribution.

Table 5: Estimation results with random DTT coefficient

Dependent variable is the location choice indicator. Estimation is based on the whole sample of 1,007,383 observations. The coefficients of $TAXD_{njt}$ and DTT_{njt} are defined as random, other coefficients are fixed. Simulation is based on 500 Halton-draws and robust standard errors are reported.

	(1)		(2)	
	MIXL		G-MNL	
	Estimate	Std. err.	Estimate	Std. err.
$TAXD_{njt}$ (mean)	-1.321***	0.251	-1.739***	0.349
$TAXD_{njt}$ (std. dev.)	3.732***	0.321	3.137***	0.309
DTT_{njt} (mean)	1.224***	0.100	2.030***	0.171
DTT_{njt} (std. dev.)	0.627***	0.198	1.039***	0.166
SUB_{njt-1}	0.933***	0.030	1.523***	0.040
TCR_{jt}	-0.165***	0.031	-0.267***	0.040
$\log(GDP)_{jt}$	0.544***	0.014	0.632***	0.020
$\log(GDPC)_{jt}$	-0.109***	0.041	-0.126**	0.055
$GDPGR_{jt}$	0.019***	0.005	0.019***	0.007
$COSTBS_{jt}$	-0.012***	0.001	-0.015***	0.002
$INFL_{jt}$	0.001	0.003	-0.004	0.005
$CORRUPT_{jt}$	-0.001	0.001	-0.001	0.002
$PRIGHT_{jt}$	0.015***	0.001	0.016***	0.002
$\log(DIST)_{nj}$	-0.599***	0.015	-0.673***	0.018
$CONTIG_{nj}$	0.089**	0.043	0.039	0.049
$COLONY_{nj}$	0.465***	0.038	0.478***	0.042
$COMLANG_{nj}$	0.060*	0.036	0.023	0.044
τ			0.686***	0.026
γ			-0.913***	0.126
AIC	71,771.80		68,972.79	
BIC	71,972.79		69,197.42	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

7.1 Postestimation statistics

Table 6 displays mean values of firm characteristics, where we distinguish broadly between two types of variables. The variables $MTOAS_n$, $MTURN_n$, $MTIAS_n$, and $MEBTA_n$ are measures of affiliate characteristics and correspond to averages over all affiliates that belong to MNE n . The variables $NLOC_n$, $NIND_n$, and $NAFF_n$ present count variables on what we might call firm scope.

Let us be more precise. $MTOAS_n$ measures the average of the total assets over all affiliates that belong to MNE n ; $MTURN_n$ measures the average of the sales over all affiliates that belong to firm n ; $MTIAS_n$ measures the average of the intangible assets; and $MEBTA_n$ measures the average of the earnings before taxes.¹⁴ The variables $NLOC_n$ and $NIND_n$ report the number of locations ($NLOC_n$) and industries ($NIND_n$) an MNE is operating in; $NAFF_n$ is the number of affiliates per MNE n .

Each row in Table 6 reports the mean of a given firm characteristic by percentile of the estimated individual-level parameter on $TAXD_n$ (taken from specification (3) in Table 2). The column denoted by (1) includes the average of the respective firm characteristic if the estimate on $TAXD_n$ lies in the first percentile of the parameter distribution (i.e., the estimated individual-level coefficient has a large negative value). Following this logic, the column denoted by (25) reports observations in the 1-to-25-percentile range.

The statistics suggest the following. It appears that firms become larger in size the less tax-responsive they are, i.e. the less negative the estimate on $TAXD_n$. For the variable $MTOAS_n$, we see that the average size roughly doubles when comparing columns (1) to (25), (50) to (75), and (75) to (99). Only the upper 1% of firms are smaller when taking $MTOAS_n$ to measure size. The pattern is not that clear when we consider $MTURN_n$ and $MEBTA_n$. However, the latter variable, earnings before taxes, remains relatively stable across the distribution. Interesting are the findings for the intangible assets $MTIAS_n$: on average, firms that do not negatively respond to taxes have a larger amount of intangible assets. This is consistent with a large literature suggesting that (i) intangible assets allow MNEs to avoid taxes (e.g., Dischinger and Riedel, 2011); (ii) preferential tax treatment for intangible assets applies (e.g., Evers *et al.*, 2015).

¹⁴Note that all variables are measured in million US dollars and we exploit the whole time span available in our data to calculate the mean.

The last three lines show that non-responding MNEs are, on average, more global and more diversified – as they are active in more countries and industries. This can be seen from the higher values of $NLOC_n$ and $NIND_n$. The interpretation of the three count variables is straightforward. For example, $NLOC_n = 7.2$, as in column (75), means that the average MNE is active in about 7 host countries. Furthermore, the average number of affiliates $NAFF_n$ of the most responsive firms is about half the one of the unresponsive ones.

The findings in Table 6 support the following conclusions. The tax-responsiveness of firms is strongly correlated with firm characteristics measuring firm size and firm scope. This is consistent with the notion that large MNEs, whose operations span across many industries and countries, are more able to avoid taxes – and therefore do not respond to statutory tax measures (see, for example, Desai *et al.*, 2006). Note that we should interpret the numbers for the very responsive firms (99-to-100-percentile range) cautiously for several reasons. One reason is that the number of observations on which the statistics are based is quite small.¹⁵ Another reason is that it can well be that the MNEs with unusually strong positive responses differ from other MNEs. Size is naturally only a proxy for the *effective tax burden* (see above), which we argue ultimately determines a firm’s tax-responsiveness (see also next section).

Table 6: Heterogenous tax-responses and firm characteristics

The table reports postestimation statistics on a number of firm characteristics along the distribution of tax parameters. The mean values for $MTOAS_n$, $MTURN_n$, $MTIAS_n$, and $MEBTA_n$ are defined in million US dollars. $NLOC_n$, $NIND_n$, and $NAFF_n$ present count variables. The estimated individual-level parameters are taken from specification (3) in Table 2.

	Percentile of heterogenous tax parameter					
	(1)	(25)	(50)	(75)	(99)	(100)
	Mean	Mean	Mean	Mean	Mean	Mean
$MTOAS_n$	69.897	140.226	147.276	227.456	301.230	72.305
$MTURN_n$	28.392	94.941	153.353	142.762	102.505	48.283
$MTIAS_n$	1.339	3.857	3.828	5.828	7.717	9.632
$MEBTA_n$	0.098	0.112	0.137	0.124	0.135	0.115
$NLOC_n$	2.074	3.303	5.406	7.239	7.307	4.423
$NIND_n$	6.037	6.580	8.723	10.432	10.297	10.654
$NAFF_n$	10.037	15.203	24.682	23.639	23.178	20.885

¹⁵While the statistics produced for the (25) to (99) percentiles are based on about 670 observations (firms), the ones for the (1) and (100) are sometimes based on less than 30 observations (firms), depending on the measured firm characteristic.

7.2 Heterogeneous tax-responses and effective tax payments

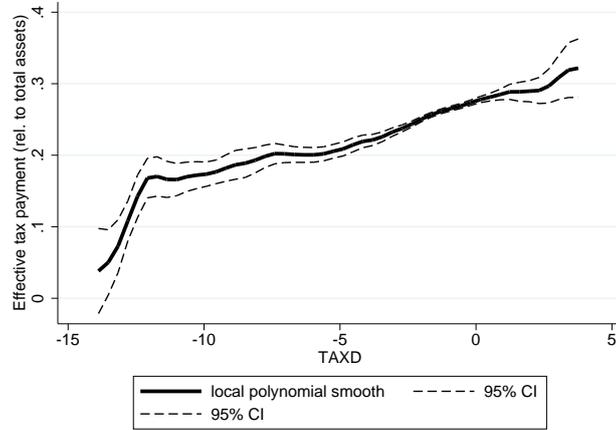
To understand the relationship between actual tax payments and estimated tax parameter, we produce Figure 3 in three steps. We first calculate the three variables depicted on the vertical axes to measure the effective tax burden of an affiliate. We define these variables as the statutory tax rate at host country j minus the effective tax rate of an affiliate, for which the MNE n has chosen location j (note that we use the index nj for variables at the level of affiliates). We calculate three versions of effective tax rates.¹⁶ One where we divide actual tax payments ($ETAX$; the information is taken from *Orbis*) by the total assets (TA ; also taken from *Orbis*) of the affiliate: $SETAX_{nj}^{TA} = TAX_j - ETAX_{nj}/TA_{nj}$; one where we divide by the operating profits (OP): $SETAX_{nj}^{OP} = TAX_j - ETAX_{nj}/OP_{nj}$; one where we divide by the earnings before tax (EBT): $SETAX_{nj}^{EBT} = TAX_j - ETAX_{nj}/EBT_{nj}$. The difference between the statutory tax rate and the effective tax rates is a proxy for the extent to which the affiliate can avoid paying tax: low values (on average, $SETAX_{nj}^{TA}$, $SETAX_{nj}^{OP}$, and $SETAX_{nj}^{EBT}$ are positive in our sample) indicate a broader tax base and that the affiliate pays a substantial amount of taxes, while high values suggest an effective tax rate, which is relatively low compared to the statutory rate. We then run kernel-weighted local polynomial regressions of the tax variables ($SETAX_{nj}^{TA}$, $SETAX_{nj}^{OP}$, $SETAX_{nj}^{EBT}$) on the estimated individual-level tax parameter (taken from specification (3) in Table 2).

In the third step, based on these local polynomial regressions, we depict the estimated relationship in Figure 3. The figure suggests the following: those firms with very negative tax-responses can avoid taxes to a lesser extent than those firms for which we have estimated a tax parameter close to zero or even positive. The less negative the tax-response, the larger the differential between statutory tax rate and effective tax payment. This implies that the mechanism producing heterogeneity in tax parameters is actually related to the extent to which firms can avoid tax payments, which supports our theoretical argument from the beginning. The dashed lines depict 95% confidence bands. For more extreme (positive and negative) values of estimates on $TAXD_n$, these become broader. For small negative values on $TAXD_n$, however, the estimates are very precise.

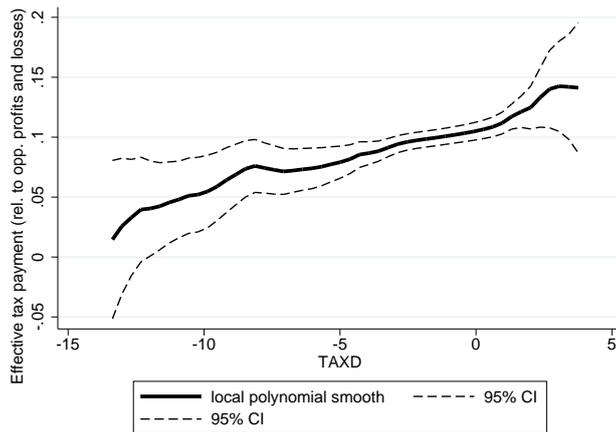
¹⁶Similar measures have been used in previous studies (e.g., Dyreng and Lindsey, 2009).

Figure 3: Effective tax payments along the tax parameter distribution

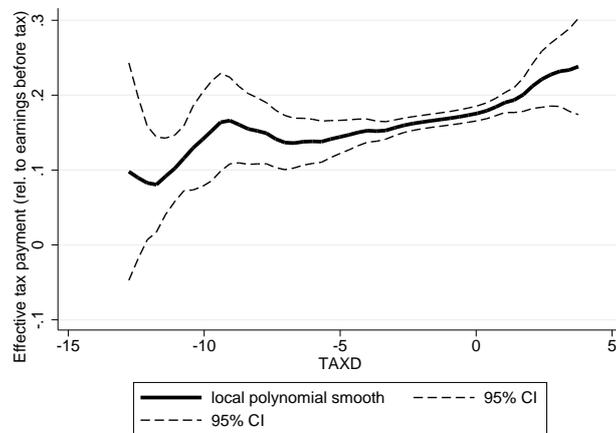
The graphs display how effective tax payments (expressed as effective tax payment divided by the total assets in (a), effective tax payment divided by operating profits in (b), and effective tax payment divided by earnings before tax in (c); variables on the vertical axes correspond to $SETAX_{nj}^{TA}$, $SETAX_{nj}^{OP}$, and $SETAX_{nj}^{EBT}$, as defined in the text) vary along the estimated individual-level parameter on $TAXD_n$ (here indicated by $TAXD$).



(a) Tax payments relative to total assets



(b) Tax payments relative to operating profits



(c) Tax payments relative to earnings before tax

8 Interaction between tax differentials and DTTs

This and the following section test Hypothesis 2 on the tax-responsiveness of MNEs' location choices in the presence of a DTT between potential host country and ultimate owner country. By considering the interaction effect between tax differentials and DTTs. Based on the finding that the G-MNL specification of the location choice model in Tables 2 and 5 provides the best model fit among all discrete choice models considered in this paper, as well as the empirical evidence for scale heterogeneity, Table 7 presents estimation results of a G-MNL specification that includes an interaction term between $TAXD_{njt}$ and DTT_{njt} . The coefficient of $TAXD-DTT_{njt}$ is estimated as a fixed parameter in the first specification and modeled as a random coefficient in the second specification. Specifically, $TAXD-DTT_{njt}$ measures the effect of an increase of the tax differential when a bilateral DTT exists between potential host country and ultimate owner country. As expected, the estimated mean coefficient has a negative sign and is statistically significant at the 5% level. However, there is no heterogeneity with respect to the interaction term as the estimated standard deviation in the second specification in Table 7 is not statistically significant. Thus, we find an unambiguously negative tax effect in the case that a DTTs exists for a given country-pair in a given year. Qualitatively, the effects of the other variables on the location choice probability do not change and all coefficients have the same signs as before. However, the estimated mean coefficient of $TAXD_{njt}$ is now only statistically significant at the 10% level, whereas the estimated standard deviation is still highly significant at the 1% level.

These results support Hypothesis 2, meaning that a DTT between potential host country and ultimate owner country does seem to make location choices more responsive to statutory corporate tax rates. In order to further investigate the effects of DTTs on the location choice probability for different tax rate differentials and to illustrate our claim of an increased tax-responsiveness of MNEs in such a scenario, we turn to a comparative static analysis in the next step.

Table 7: Estimation results with random $TAXD$ - DTT coefficient

Dependent variable is the location choice indicator. Estimation is based on the whole sample of 1,007,383 observations. The coefficient of $TAXD_{njt}$ is defined as random in specification (1), in specification (2) the coefficient of $TAXD$ - DTT_{njt} is also defined as random. Simulation is based on 500 Halton-draws and robust standard errors are reported. The simulated log-likelihood at convergence (LL) is reported as a positive number.

	(1)		(2)	
	G-MNL		G-MNL	
	Estimate	Std. err.	Estimate	Std. err.
$TAXD_{njt}$ (mean)	-0.952*	0.514	-0.962*	0.515
$TAXD_{njt}$ (std. dev.)	1.978***	0.466	1.957***	0.467
$TAXD$ - DTT_{njt} (mean)	-0.912**	0.394	-0.903**	0.396
$TAXD$ - DTT_{njt} (std. dev.)			-0.143	0.156
DTT_{njt}	1.486***	0.088	1.487***	0.088
SUB_{njt-1}	1.510***	0.039	1.510***	0.039
TCR_{jt}	-0.279***	0.040	-0.279***	0.040
$\log(GDP)_{jt}$	0.635***	0.020	0.635***	0.020
$\log(GDPC)_{jt}$	-0.122**	0.055	-0.122**	0.055
$GDPGR_{jt}$	0.018***	0.007	0.018***	0.007
$COSTBS_{jt}$	-0.016***	0.002	-0.016***	0.002
$INFL_{jt}$	-0.002	0.005	-0.002	0.005
$CORRUPT_{jt}$	-0.001	0.002	-0.001	0.002
$PRIGHT_{jt}$	0.016***	0.002	0.016***	0.002
$\log(DIST)_{nj}$	-0.675***	0.018	-0.675***	0.018
$CONTIG_{nj}$	0.039	0.049	0.039	0.049
$COLONY_{nj}$	0.488***	0.041	0.488***	0.041
$COMLANG_{nj}$	0.024	0.043	0.024	0.043
τ	0.674***	0.025	0.674***	0.025
γ	-3.368***	0.985	-3.405***	1.001
LL	34,479.04		34,478.98	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

9 The consequences of DTTs

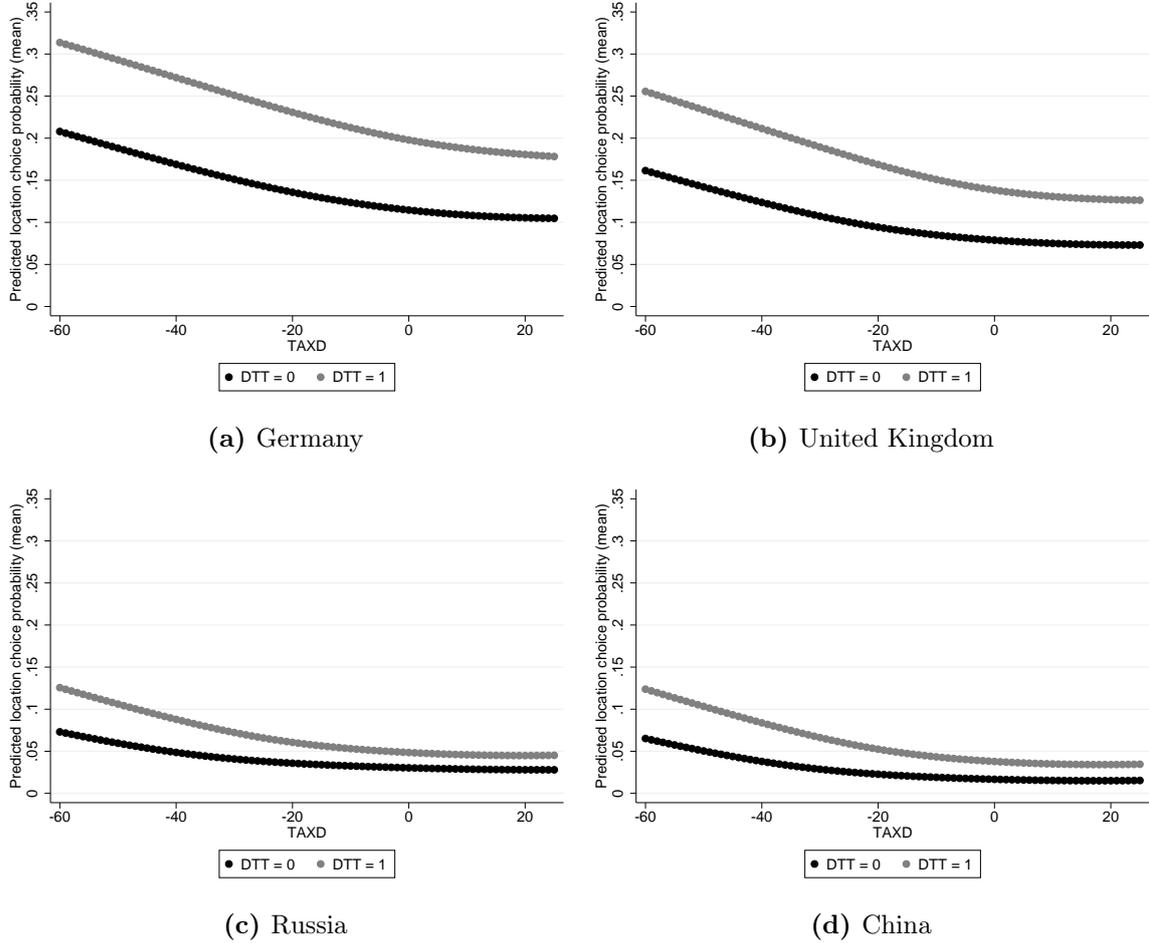
We are interested in the question of how the predicted location choice probability changes for different statutory corporate tax rate differentials, conditional on the (non-)existence of a bilateral DTT between potential host country and ultimate owner country. This is done in a comparative static analysis, where we calculate the average predicted location choice probability for different host countries over a range of tax rate differentials and by distinguishing between two scenarios of (i) an existing bilateral DTT and (ii) no existing bilateral DTT.¹⁷ Figure 4 plots the average predicted probability of a firm setting up an affiliate in a host country as a function of the tax differential between the host country's and the ultimate owner country's statutory corporate tax rate for, as in Section 7, Germany and the United Kingdom as well as Russia and China.

Several interesting findings can be derived from the plots. While a negative tax rate differential increases the predicted location choice probability for all host countries and both scenarios, the marginal effect of a DTT, i.e. the difference between the two curves, decreases for an increasing tax rate differential (the tax rate in the host country increases, relative to the tax rate in the ultimate owner country). The probability that an MNE establishes a foreign affiliate in a host country that has concluded a bilateral DTT with the ultimate owner country is greater for locations with a lower tax rate, compared to the ultimate owner country, which is in line with Hypothesis 2. This effect is even more pronounced for Russia and China, both countries with a lower base probability compared to Germany and the United Kingdom. Thus, firms seem to react more responsive to taxes in such countries, whereas countries like Germany and the United Kingdom might be endowed with other location attributes that render the responsiveness to taxes less pronounced than for countries like Russia and China.

¹⁷In our data, the values for $TAXD_{njt}$ range from -55% to 42% (see Table 1). For illustrative purposes, we calculate average predicted location choice probabilities for tax differentials up to -60%, which should be considered an extreme case. On the other hand, we have chosen the upper cut-off point to be 25%, since only for Germany and China there are some observations in the choice set with values of $TAXD_{njt}$ exceeding 25%. In all these cases, the MNE making the location choice is located in a tax haven.

Figure 4: Effects of $TAXD$ and DTT on the predicted location choice probability

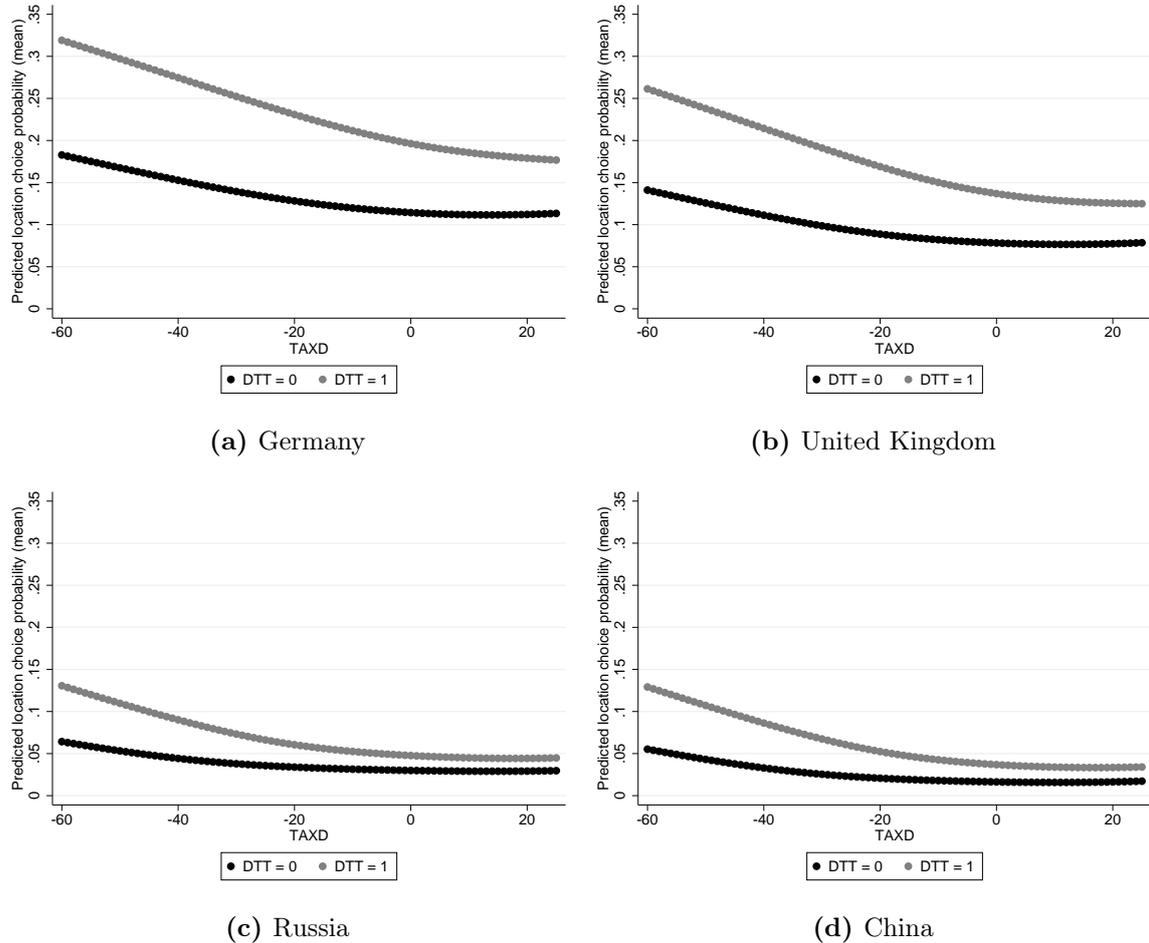
The figure depicts predicted average location choice probabilities across different statutory tax rate differentials for four potential host countries, (a) Germany (b) United Kingdom, (c) Russia, and (d) China. Two scenarios are illustrated: an existing DTT ($DTT = 1$) between potential host country and ultimate owner country as well as the absence of such a bilateral DTT ($DTT = 0$). Calculations are based on specification (3) in Table 2 and 500 Halton-draws.



In addition, referring to the previous section, we consider the effect of the interaction term $TAXD \cdot DTT_{njt}$ in the same comparative static framework and for the same host countries. Figure 5 illustrates the findings, based on specification (1) in Table 7. Compared to Figure 4, the marginal effect of a DTT is slightly larger. This becomes apparent as one moves along the horizontal axis and the tax differential becomes more negative. The two curves diverge even more. For example, the maximum difference in average predicted probabilities for the United Kingdom is now 12%, instead of 9% in Figure 4. Thus, the effect stated earlier becomes more pronounced if the interaction effect between tax rate differentials and DTTs is taken into account.

Figure 5: Effects of $TAXD$, DTT , and $TAXD-DTT$ on the predicted location choice probability

The figure depicts predicted average location choice probabilities across different statutory tax rate differentials for four potential host countries, (a) Germany (b) United Kingdom, (c) Russia, and (d) China. Two scenarios are illustrated: an existing DTT ($DTT = 1$) between potential host country and ultimate owner country as well as the absence of such a bilateral DTT ($DTT = 0$). Moreover, for these four host countries and both scenarios the effect of $TAXD-DTT$ is also taken into account. Calculations are based on specification (1) in Table 7 and 500 Halton-draws.



10 Concluding remarks

This paper provides new evidence on the consequences of DTTs on the location choices of MNEs. Our analysis accounts for interactions between DTTs and statutory tax rate differentials and we particularly allow the tax rate parameters to vary randomly across MNEs. We provide a formal argument for why the latter should be the case if firms differ in ability to avoid taxes. Given the finding of substantial heterogeneity in tax-responses, we examine the potential drivers behind heterogeneous tax-responses and what this means for the implementation of DTTs. We also present postestimation statistics along the distribution of the tax

parameters to show that our findings are consistent with the argument that tax avoidance determines the response to statutory tax rules.

Our results suggest that increasing tax differentials between potential host countries and the country of the ultimate owner have a negative effect on location choice probability. While DTTs have a positive effect, the precise magnitude in terms of change in the probability to choose a location is specific to a country. We predict the consequences of DTTs on location choices and estimate location choice response functions to tax rate differentials. We illustrate the effects by focusing on four countries which are frequently chosen as a foreign location in our sample: Germany, the United Kingdom, Russia, and China. We then graphically demonstrate that the marginal tax effect becomes larger the more negative the tax differential becomes, i.e. the lower the foreign tax is relative to the tax rate at the location of the ultimate owner. This is plausible, as tax considerations for firms where the ultimate owner resides in a low-tax country are probably not too important. If the tax differential becomes zero or even positive, the marginal tax effect goes to zero and the tax-response function becomes flat. The same pattern is found for the impact of a DTT: the more negative the tax differential, the larger the effect of concluding a DTT. If the tax differential between ultimate owner and foreign country is zero or positive, the positive effect of a DTT becomes very small.

We finally test whether countries successfully restrict tax avoidance by implementing DTTs. These often include information exchange or other rules to prevent MNEs from profit shifting. We show that the tax elasticity becomes larger if a DTT is signed between two countries. This is consistent with the notion that a central intention of governments is to reduce the scope for tax avoidance of MNEs by signing a DTT.

References

- Allenby, G. M. and Rossi, P. E. (1998) Market models of consumer heterogeneity, *Journal of Econometrics*, **89**, 57–78.
- Arulampalam, W., Devereux, M. P. and Maffini, G. (2017) Taxes and the location of targets, Working Paper.
- Bloningen, B. A. and Davies, R. B. (2004) The effects of bilateral tax treaties on U.S. FDI activity, *International Tax and Public Finance*, **11**, 601–622.
- Davies, R. B., Norbäck, P.-J. and Tekin-Koru, A. (2009) The effect of tax treaties on multinational firms: New evidence from microdata, *The World Economy*, **32**, 77–110.
- Desai, M. A., Foley, C. F. and Hines, J. R. (2006) The demand for tax haven operations, *Journal of Public Economics*, **90**, 513–531.
- Devereux, M. P. and Griffith, R. (1998) Taxes and the location of production: Evidence from a panel of US multinationals, *Journal of Public Economics*, **68**, 335–367.
- Devereux, M. P. and Maffini, G. (2007) The impact of taxation on the location of capital, firms and profit. A survey of empirical evidence, Oxford University Centre for Business Taxation Working Paper No. 10/03.
- Dischinger, M. and Riedel, N. (2011) Corporate taxes and the location of intangible assets within multinational firms, *Journal of Public Economics*, **95**, 691–707.
- Dyreng, S. D. and Lindsey, B. P. (2009) Using financial accounting data to examine the effect of foreign operations located in tax havens and other countries on U.S. multinational firms' tax rates, *Journal of Accounting Research*, **47**, 1283–1316.
- Egger, P. H., Larch, M., Pfaffermayr, M. and Winner, H. (2006) The impact of endogenous tax treaties on foreign direct investment: Theory and evidence, *Canadian Journal of Economics*, **39**, 901–931.
- Egger, P. H. and Merlo, V. (2011) Statutory corporate tax rates and double-taxation treaties as determinants of multinational firm activity, *FinanzArchiv: Public Finance Analysis*, **67**, 145–170.

- Egger, P. H., Merlo, V. and Wamser, G. (2014) Unobserved tax avoidance and the tax elasticity of FDI, *Journal of Economic Behavior & Organization*, **108**, 1–18.
- Egger, P. H. and Stimmelmayer, M. (2017) Taxation and the multinational firm, CESifo Working Paper Series No. 6384.
- Evers, L., Miller, H. and Spengel, C. (2015) Intellectual property box regimes: Effective tax rates and tax policy considerations, *International Tax and Public Finance*, **22**, 502–530.
- Fiebig, D. G., Keane, M. P., Louviere, J. J. and Wasi, N. (2010) The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity, *Marketing Science*, **29**, 393–421.
- Griffith, R., Miller, H. and O’Connell, M. (2014) Ownership of intellectual property and corporate taxation, *Journal of Public Economics*, **112**, 12–23.
- Holmes, K. (2014) *International Tax Policy and Double Tax Treaties: An Introduction to Principles and Application*, IBFD Publications BV.
- Keane, M. P. and Wasi, N. (2013) Comparing alternative models of heterogeneity in consumer choice behavior, *Journal of Applied Econometrics*, **28**, 1018–1045.
- Louviere, J. J., Meyer, R. J., Bunch, D. S., Carson, R., Dellaert, B., Hanemann, W. M., Hensher, D. and Irwin, J. (1999) Combining sources of preference data for modelling complex decision processes, *Marketing Letters*, **10**, 205–217.
- Louviere, J. J., Street, D., Carson, R., Ainslie, A., Deshazo, J. R., Cameron, T., Hensher, D., Kohn, R. and Marley, T. (2002) Dissecting the random component of utility, *Marketing Letters*, **13**, 177–193.
- Merlo, V., Riedel, N. and Wamser, G. (2016) The impact of thin capitalization rules on the location of multinational firms’ foreign affiliates, Research School of International Taxation Working Paper 01/2016.
- Revelt, D. and Train, K. E. (2000) Customer-specific taste parameters and mixed logit: Households’ choice of electricity supplier, Working Paper No. E00-274.
- Train, K. E. (2009) *Discrete choice methods with simulation*, Cambridge University Press, second edn.

Appendix A

Table A.1 lists the various sources from which data were collected. The construction of the dataset used in our empirical analysis involved several steps. A detailed description is available upon request.

Table A.1: Data sources

Data Source	Data on
Bureau van Dijk, Orbis	Location of MNEs and affiliates
CEPII	Measures for distance, contiguity, colonial relationship, common language
Heritage Foundation, Heritage Indicators	Freedom from corruption index, property rights index
IMF, World Economic Outlook	Consumer price changes
International Bureau of Fiscal Documentation; Tax surveys by Ernst&Young, PwC, KPMG	Statutory corporate tax rates, thin-capitalization rules, and double taxation treaties
World Bank, World Development Indicators	GDP, GDP per capita, GDP growth, domestic credit to private sector, cost of business start-up procedures