

Rechbauer, Martina

Working Paper

Identifying firms' tax loss carry-forward status: The accuracy of database-ariven methods

arqus Discussion Paper, No. 201

Provided in Cooperation with:

arqus - Working Group in Quantitative Tax Research

Suggested Citation: Rechbauer, Martina (2016) : Identifying firms' tax loss carry-forward status: The accuracy of database-ariven methods, arqus Discussion Paper, No. 201, Arbeitskreis Quantitative Steuerlehre (arqus), Berlin

This Version is available at:

<https://hdl.handle.net/10419/129105>

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.



Arbeitskreis Quantitative Steuerlehre
Quantitative Research in Taxation – Discussion Papers

Martina Rechbauer

**Identifying Firms' Tax Loss Carry-Forward Status
- The Accuracy of Database-Driven Methods -**

arqus Discussion Paper No. 201

January 2016

www.arqus.info

ISSN 1861-8944

Martina Rechbauer*

Identifying Firms' Tax Loss Carry-Forward Status

– The Accuracy of Database-Driven Methods –

Abstract: Due to data restrictions, empirical tax research commonly relies on database-driven methods as a means of identifying firms' tax loss carry-forward (TLCF) status. Employing a panel of listed Italian parent companies, I am the first to empirically examine the accuracy of database-driven methods in predicting the availability and the amount of TLCF at single-firm level. In order to assess the accuracy of database-driven identification methods, I compare firms' true TLCF status, as determined based on IFRS statement information, to the TLCF status predictions of the methods examined. I find that database-driven methods do not perform well in predicting the availability of TLCF. They perform poorly in predicting the amount of TLCF available to firms. Empirical studies that rely on database-driven identification methods might thus not be able to derive reliable results regarding the impact of TLCF. My findings thus indicate that there is a strong need for firm-specific TLCF information provided by local authorities in empirical tax research.

Keywords: identification, tax loss carry-forwards, database-driven methods

JEL Classification: C81, H25, K34

* Martina Rechbauer, Research Assistant, Institute of Accounting and Taxation, University of Graz, Universitätsstraße 15/G2, 8010 Graz, Austria, e-mail: martina.rechbauer@uni-graz.at. I thank Rainer Niemann and my colleagues from the University of Graz for their constructive feedback. I also highly appreciate helpful comments from the participants of the 2014 WU DIBT Seminar in Vienna, the 2014 EAA Annual Congress in Tallinn and the 2014 VHB Annual Congress in Leipzig. Financial support by the Austrian Science Fund (FWF) [P 22324-G1] is gratefully acknowledged.

1 Introduction

Due to fiscal secret, information on the amount of tax loss carry-forwards (TLCF) available to firms is typically not available to empirical tax research. For this reason, studies that empirically examine how TLCF affect firm behavior commonly rely on database-driven methods as a means of identifying firms' TLCF status. For example, Overesch and Voeller (2010) investigate how TLCF influence the financing behavior of firms. They assume that firms have TLCF if their earnings before taxes (EBT) were negative in the year before. In a similar setting, Bernasconi, Marenzi and Pagani (2005) expect firms to have TLCF if their EBT was negative in the two preceding years. Dreßler and Overesch (2013) examine how TLCF influence the investment behavior of firms. They expect firms with loss carry-forwards in terms of commercial law to be exposed to TLCF too. In order to obtain reliable results regarding the impact of TLCF, database-driven methods should be accurate in identifying the TLCF status of firms. This paper empirically investigates whether or not this is true.

I examine the accuracy of database-driven methods in predicting both the availability and the amount of TLCF at single-firm level. My analysis is based on a panel of listed Italian parent companies between 2010 and 2012 (325 firm-year observations). Employing a panel of listed Italian parent companies proves to be essential for the aim of this study. Listed Italian firms have been obliged to prepare their unconsolidated financial statements in accordance with the IFRS since 2006. The provisions imposed by IAS 12.81 require that firms disclose information on their TLCF status in the notes on deferred taxes. This allows me to determine firms' true TLCF status in a given year and hence, to examine the accuracy of database-driven identification methods by comparing firms' true TLCF status to the TLCF status predictions of the methods examined. The methods I examine are loss carry-forwards based on IFRS earnings and pre-tax cashflow realized in the past and methods based on industry affiliation.

In order to assess the accuracy of database-driven methods in predicting the availability of TLCF, I determine each method's percentage of correct TLCF status predictions made. I find that none of the methods examined performs explicitly well in identifying firms' true TLCF status. For the most accurate method, a loss carry-forward based on firms' EBT from four past years, the percentage of correct predictions made is equal to 79.08% only. The method's accuracy is likely to decrease even further in settings where a large number of firms operating in the industries Capital Goods, Consumer Durables & Apparel or Media (four-digit Global Industrial Classification Standard (GICS) codes 2010, 2520 and 2540, respectively) is employed. In order to assess the accuracy of database-driven

methods in predicting the amount of TLCHF available, I determine each method's percentage of small logarithmic prediction errors (where firms' true TLCHF amount is either not more than 1.11 or not more than 0.9 times as large as the predicted amount). I find that the methods examined perform quite poorly. For the most accurate method, a loss carry-forward based on firms' earnings before interest and taxes (EBIT) from three past years, the percentage of small prediction errors is equal to 19.64% only. Furthermore, the method's accuracy varies substantially by year of observation.

Overall, my results suggest that empirical studies relying on database-driven methods similar to the ones tested here might not be able to properly identify firms' TLCHF status. Any results obtained regarding the impact of TLCHF are thus likely to be distorted. Previous studies that are likely to suffer from these issues include the investigations of Bernasconi et al. (2005), Overesch and Voeller (2010), Overesch and Wamser (2010), Buettner, Overesch and Wamser (2011), Buettner, Overesch, Schreiber and Wamser (2012), Dreßler and Overesch (2013), Stöckl and Winner (2013) and Overesch and Wamser (2014). In order to verify the accuracy of the results obtained in these studies, I recommend conducting replication studies. These studies should rely on firms' true TLCHF status instead of database-driven proxies. For future studies, my findings indicate that database-driven identification methods should not be used at all. In empirical tax research, there is thus a strong need for firm-specific TLCHF status information provided by local authorities.

My analysis adds to the scarce literature on the accuracy of database-driven identification methods. Although database-driven methods are used extensively in empirical tax research, only Kinney and Swanson (1993), Mills, Newberry and Novack (2003) and Niemann and Rechbauer (2013) have empirically investigated their accuracy so far. Kinney and Swanson (1993) and Mills et al. (2003) examine the accuracy of Compustat's data item # 52. This data item is typically employed by empirical tax research in the U.S. in order to identify firms' true TLCHF status (see, for example, Mackie-Mason, 1990, and Barclay and Smith, 1995). Niemann and Rechbauer (2013) investigate the accuracy of two effective tax rate measures, a measure based on the existence of a negative EBT in the three years prior to the observation period and a measure based on the existence of a negative EBIT in the year before. All three studies conclude that database-driven methods do not perform well in identifying firms' TLCHF status.

Empirical tax research typically examines how TLCHF affect the behavior of firms at single-firm level. Due to data restrictions, the latter studies, however, assess the methods' accuracy at group level only. For many empirical studies, they thus do not provide valuable insights. My analysis is the first to provide empirical evidence on the accuracy of

database-driven identification methods at single-firm level and thus, offers insights highly relevant for empirical tax research. I examine only methods that are based on standard data items, which are typically available in financial databases. This ensures that my results are relevant for studies using data from various data sources. Moreover, due to the fact that I investigate the accuracy of loss carry-forwards based on IFRS earnings, my findings tend to be highly relevant for studies that examine the impact of TDCF in countries that either prescribe or allow the appliance of the IFRS,¹ given that their institutional setting is similar to that of Italy. The same applies to studies in countries with accounting standards similar to the IFRS.

The remainder of this paper is organized as follows. Section 2 outlines the institutional setting of my analysis. Section 3 explains the way I determine firms' true TDCF status and each method's TDCF status prediction. Moreover, it provides insights into how I assess the methods' accuracy. Information on the sample employed is given in Section 4. Section 5 discusses the results derived. Section 6 concludes.

2 Institutional Setting

Italian firms are subject to a corporate income tax (*Imposta sul Reddito delle Società*, IRES), which is governed by the country's consolidated income tax text.² Firms' taxable income is taxed at a constant rate of 27.5%, which has been applied since 2008. Previously, it was equal to 33%. Firms that carry on certain activities in the fields of energy production and supply are subject to an increased rate of 34%. The increased corporate income tax rate applies if previous revenue exceeds a level of 3,000 thousand euros (TEUR) and previous taxable income exceeds a level of 300 TEUR. The increased rate was introduced in 2009 at a level of 33%. It had to be applied if revenue exceeded 25,000 TEUR. Between 2011 and 2013, the rate was increased to 38% with threshold levels for revenue and taxable income of 10,000 and 1,000 TEUR, respectively. There is no minimum tax for listed firms.

For IRES purposes, Italian firms are allowed to carry forward negative taxable income

¹ Besides Italy, the following European countries require that listed firms follow the IFRS at single-firm level: Albania, Bulgaria, Cyprus, the Czech Republic, Estonia, Georgia, Greece, Iceland, Lithuania, Macedonia, Malta, Montenegro, Russia, Serbia and the Slovak Republic. At single-firm level, IFRS can be applied in Bosnia and Herzegovina, Denmark, Finland, Greenland, Ireland, Luxembourg, the Netherlands, Norway, Poland, Slovenia and the United Kingdom.

² Listed Italian firms are also subject to a regional tax on productive activities (*L'Imposta Regionale sulle Attività Produttive*, IRAP) of basically 3.5% (3.9% until 2013). The IRAP is governed by Decreto Legislativo 466/1997. It does not offer any loss carry-back or loss carry-forward provisions.

to subsequent years in order to reduce future taxable income. Tax losses incurred before 2011 can be carried forward for five years with no deduction limit. Tax losses incurred in 2011 or later can be carried forward indefinitely. However, the TLCF deduction for these losses is limited to an amount of 80% of firms' positive taxable income. Italy does not offer any loss carry-back provisions.

Listed Italian parent companies have to prepare their unconsolidated financial statements in accordance with the IFRS. Taxable income of these firms is derived by adjusting their IFRS result in order to meet tax provisions that are different from the IFRS. In 2008, this derivation principle was reinforced. Since then, IFRS rules governing income qualification, timing of computation and classification have become fully relevant for tax purposes, even in cases where tax and IFRS provisions are not in line with one another. In contrast to that, opposing tax provisions concerning income evaluation and quantification have remained relevant. These rules govern, amongst others, the deductibility of interest expense, the taxation of dividend income and capital gains/losses from equity investments, and the depreciation of tangible assets (Giacometti, 2009). From 2008 on, interest expense exceeding interest income can only be deducted up to an amount of 30% of firms' earnings before interest, taxes, depreciation and amortization (EBITDA). Interest expense not deductible within a year can be carried forward to subsequent years. The portion of EBITDA not used within a year may also be carried forward. Previously, a thin-capitalization rule based on a safe debt-to-equity ratio of 4:1 restricted the deductibility of interest paid to qualified shareholders holding at least 25% of the firm's shares. Additional limitations were imposed by equity pro rata rules. In firms' IFRS result, interest expense is fully recognized. Dividend income from equity investments is tax-exempt for 95% of the total amount, given that the investment is classified as a non-current financial asset, and that the participated firm is not located within a tax haven country. Capital gains on disposals of equity investments are also tax-exempt for 95% of the total amount if the following conditions are met: the investment was held for at least twelve months and classified as a non-current financial asset, the participated firm is not located within a tax haven country and engages in commercial activities. The tax-exempt proportion of capital gains was set equal to 95% in 2008. Previously, it varied between 84% and 100%. Capital losses on equity investments that qualify for the participation exemption are not tax deductible. In their IFRS result, firms' dividend income as well as their capital gains/losses from equity investments are fully recognized. From 2008 onwards, depreciation of tangible assets has been permitted on a straight-line basis only. Previously, it was possible to double depreciation allowances in the first three years of an asset's life, or to increase the amount of depreciation allowance in case of intensive utilization. The IFRS

allow various depreciation methods to be applied.

For IRES purposes, Italian firms can opt for being part of a group taxation regime. Under this regime, taxable income of group members is either fully (domestic members) or partially (foreign members) attributed to the group's parent company. The taxable income for the entire group is then taxed at parent company level only.

3 Research Design

In order to assess the accuracy of database-driven identification methods, I compare firms' true TDCF status to the TDCF status predictions of the methods examined. The following two subsections provide insights into how I derive firms' true TDCF status and each method's TDCF status prediction. Information on the measures used to make judgements regarding the methods' accuracy is given in the last subsection.

3.1 Derivation of Firms' True TDCF Status

I determine firms' true TDCF status by relying on information on TDCF published in the notes to firms' unconsolidated IFRS statement. According to the IFRS, firms are not obliged to disclose their total stock of TDCF in the notes to their financial statement. However, under the terms of IAS 12.81, they should publish the amount of deferred tax assets on TDCF, which are recognized in the statement of financial position, as well as the amount of TDCF for which no deferred tax assets have been recorded. This provision allows me to determine firms' true TDCF status if the stock of TDCF is not voluntarily disclosed.

An approach to determine firms' true TDCF status based on the provisions imposed by IAS 12.81 was developed by Kager, Niemann and Schanz (2011). I follow their approach by applying the formula below:

$$TAMT_{i,t-1} = \frac{DTA_{i,t-1}}{\tau_{i,t-1}} + TAMT_{i,t-1}^{NDTA} \quad (1)$$

In Formula 1, $TAMT_{i,t-1}$ is firm i 's total stock of TDCF at the end of year $t - 1$, which is available for deduction in year t . $DTA_{i,t-1}$ and $TAMT_{i,t-1}^{NDTA}$ represent the amount of deferred tax assets on TDCF recognized in the statement of financial position of year $t - 1$, and the stock of TDCF for which no deferred tax assets have been recorded, respectively. $\tau_{i,t-1}$ is the corporate income tax rate used to determine $DTA_{i,t-1}$. TDCF of Italian parent

companies can be offset against taxable income subject to the IRES only. As a result, $\tau_{i,t-1}$ corresponds to the IRES rate of firm i in the year in which it expects its TLCF to be deducted. Italian firms that do not belong to the industries Energy (four-digit GICS code 1010) and Utilities (four-digit GICS code 5510) can only be subject to the general IRES rate of 27.5% (see Section 2). For all of these firms, I assume that $\tau_{i,t-1} = 27.5\%$. Firms that belong to the industries Energy or Utilities can either be subject to the general or the increased IRES rate, depending upon whether or not their revenue, or both their revenue and their taxable income exceed the threshold levels imposed by the Italian tax authority. For these firms, I am not able to make a general assumption regarding their expected future tax rate. As a result, I am able to derive $\tau_{i,t-1}$ only if firms explicitly disclose the corporate income tax rate used to determine $DTA_{i,t-1}$. All other firms are dropped from my sample (see Section 4.1). For my further analysis, I employ two variables that represent firms' true TLCF status. Both of them relate to the year in which TLCF are available for deduction. $TAVB_{i,t}$ is a dummy variable equal to one if firm i has TLCF for deduction in year t and zero otherwise. I set $TAVB_{i,t} = 1$ if $TAMT_{i,t-1} > 0$. $TAMTD_{i,t}$ represents firm i 's amount of TLCF available for deduction in year t . It is set equal to $TAMT_{i,t-1}$.

Applying Formula 1 yields reliable TLCF estimates if firms fully follow the provisions imposed by IAS 12.81. However, if they do not, it is not possible to exactly determine firms' true TLCF status in the way described above. For my analysis, I assume that firms fully follow IAS 12.81. This is because I am unable to clearly distinguish between firms that do and firms that do not by solely examining firms' IFRS statements. Kager and Niemann (2013), who empirically examine the disclosure behavior of listed Austrian, German and Dutch firms between 2004 and 2007, raise concerns about the accuracy of this assumption. They find that firms often do not publish $TAMT_{i,t-1}^{NDTA}$ and thus, conclude that TLCF estimates based on IAS 12.81 information are likely to be distorted. In order to verify the accuracy of the above assumption regarding firms' reporting behavior, I ask my sample firms to reveal their true TLCF status during the observation period using a questionnaire. This approach allows me to correct my TLCF estimates if necessary. Moreover, it provides insights into the accuracy of determining firms' TLCF status based on IAS 12.81 information (see Section 4.2).

3.2 Derivation of Firms' Predicted TLCF Status

For each of the methods examined, I employ two variables that represent firms' predicted TLCF status. Again, both variables relate to the year in which TLCF are available for

deduction. $PAVB_{i,t}^m$ is a dummy variable equal to one if method m predicts the existence of a TLCF in year t and zero otherwise. $PAMTD_{i,t}^m$ represents firm i 's amount of TLCF available for deduction in year t , as predicted by method m .

All data necessary to compute the TLCF status predictions of the methods examined is obtained from the Amadeus database. A summary of the Amadeus data items used can be found in Supplemental Appendix A.1. Due to the fact that I investigate the accuracy of database-driven identification methods at single-firm level, I rely on unconsolidated data only.

3.2.1 Loss Carry-Forwards based on IFRS Earnings

I examine the accuracy of loss carry-forwards based on IFRS earnings (ILCF). This is because empirical tax research commonly uses loss carry-forwards in terms of commercial law as a proxy for firms' true TLCF status. For example, Buettner et al. (2012) and Dreßler and Overesch (2013) assume that firms with loss carry-forwards in terms of commercial law are also exposed to TLCF. The latter studies make use of a specific data item provided by the German MiDi database, which represents firms' amount of loss carry-forwards in terms of commercial law. If loss carry-forwards in terms of commercial law are not available as separate data items, loss carry-forward measures are constructed based on earnings realized in the past. For example, Overesch and Voeller (2010) expect firms to have TLCF if their EBT was negative in the year before. Bernasconi et al. (2005) consider firms' EBT from two past years in order to identify firms' TLCF status. The latter studies make use of the Amadeus and Aida database, respectively.

Due to fact that the Amadeus database does not offer a specific data item that represents the amount of ILCF available to firms, I construct ILCF using earnings information from past periods. In particular, I determine ILCF based on earnings from one, two, three and four past years. $ILCFAMT_{i,t-1}^n$ is defined as the amount of ILCF available to firm i at the end of year $t - 1$. It is based on earnings information from n past years (with $1 \leq n \leq 4$). I assume that firm i is not exposed to any ILCF prior to year $t - n$. Based on this assumption, $ILCFAMT_{i,t-1}^n$ is built up recursively, taking the level of earnings realized in each period between years $t - n$ and $t - 1$ into account. A mathematical derivation of $ILCFAMT_{i,t-1}^n$ is provided in Supplemental Appendix A.2. For my further analysis, I set $PAVB_{i,t}^{ILCF^n} = 1$ and hence, assume that TLCF are available for deduction in year t , if $ILCFAMT_{i,t-1}^n > 0$. If $ILCFAMT_{i,t-1}^n = 0$, I assume that no TLCF are available. $PAMTD_{i,t}^{ILCF^n}$ is set equal to $ILCFAMT_{i,t-1}^n$.

Examining the accuracy of ILCF based on earnings from one, two, three and four past

years, allows me to analyze how the accuracy of ILCF changes if the degree of past information considered is increased. I assume that the inclusion of more relevant earnings information enhances the accuracy of ILCF in identifying firms' TLCF status. However, I also expect book-tax differences to increase if ILCF are based on more past earnings information. This can have a negative effect on the method's accuracy. The overall effect of an increase in the degree of past earnings information considered depends upon which one of the latter two effects tends to be stronger. Given my setting, constructing ILCF based on earnings from five past years would also be appropriate (see Section 2). Unfortunately, I am unable to follow this approach. For listed Italian parent companies, IFRS data is available only from 2006 on. Hence, I am not able to determine five-year ILCF for my firm-year observations in 2010. Constructing ILCF based on earnings from more than five past years would not be appropriate in my setting.

I also consider different types of earnings measures in constructing ILCF. This allows me to check whether or not earnings measures that exclude income possibly subject to book-tax differences perform better in identifying firms' TLCF status than earnings measures that do not. The first measure I employ is firms' EBT. It includes all income that is possibly subject to book-tax differences. I also determine ILCF based on firms' EBIT. This measure excludes financial income. Thus, book-tax differences in firms' financial income do not negatively affect the measure's accuracy. It does, however, suffer from the fact that it also excludes those parts of financial income, which are relevant for tax purposes. I assume that firms' EBIT outperforms firms' EBT if the negative accuracy effect due to the inclusion of financial income in firms' EBT outweighs the negative effect due to the non-inclusion of financial income in firms' EBIT. The third measure I employ is firms' EBITDA. It excludes financial income and depreciation allowances. Book-tax differences in firms' financial income and depreciation allowances do not negatively affect the accuracy of this measure. It does, however, suffer from the fact that it also excludes those parts of financial income and depreciation allowances, which are relevant for tax purposes. I assume that firms' EBITDA outperforms firms' EBT if the negative accuracy effect due to the inclusion of financial income and depreciation allowances in firms' EBT is larger than the negative effect due to the non-inclusion of financial income and depreciation allowances in firms' EBITDA. The Amadeus database does not offer any other earnings measures to examine. I thus cannot address accuracy issues related to any other book-tax differences. Moreover, due to relying on unconsolidated data, I am not able to address any accuracy issues that probably arise due to firms' participation in the Italian group taxation regime.

3.2.2 Loss Carry-Forwards based on Pre-Tax Cashflow

Besides ILCF, I examine the accuracy of loss carry-forwards based on a measure of firms' pre-tax cashflow (CLCF). As for ILCF, I determine firms' CLCF by relying on cashflow information from one, two, three and four past years. $CLCFAMT_{i,t-1}^n$ is defined as the amount of CLCF available to firm i at the end of year $t-1$. It is based on pre-tax cashflow information from n past years (with $1 \leq n \leq 4$). In order to determine $CLCF_AMT_{i,t-1}^n$, I apply the same principles as for $ILCF_AMT_{i,t-1}^n$, using a measure of pre-tax cashflow instead of IFRS earnings. The cashflow data item provided by the Amadeus database includes firms' total tax expense. TLCF are, however, defined on a pre-tax basis. In order to account for this fact, I adjust the Amadeus data item by re-adding firms' total tax expense. Moreover, I do not consider firms' extraordinary income, which is also part of the Amadeus cashflow data item. This seems to be reasonable, given that firms' extraordinary income is likely to contain various non-cash items. Overall, the measure I rely on tends to be a weak measure of pre-tax cashflow. There are no additional cash items added to, and the amount of depreciation allowances is the only non-cash item excluded from firms' EBT. For my further analysis, I set $PAVB_{i,t}^{CLCF^n} = 1$ if $CLCFAMT_{i,t-1}^n > 0$. If $CLCFAMT_{i,t-1}^n = 0$, I assume that no TLCF are available. $PAMTD_{i,t}^{CLCF^n}$ is set equal to $CLCFAMT_{i,t-1}^n$.

As for ILCF, examining the accuracy of CLCF based on pre-tax cashflow from one, two, three and four past years, enables me to evaluate how the accuracy of CLCF changes if the degree of past information considered is increased. My assumptions regarding the impact of an increase in the degree of past cashflow information considered remain the same as before. Examining the accuracy of CLCF also allows me to compare the performances of CLCF and ILCF based on firms' EBT. My measure of pre-tax cashflow excludes firms' depreciation allowances, whereas firms' EBT does not. Accordingly, I assume that CLCF outperform ILCF based on firms' EBT if the negative accuracy effect due to the inclusion of the non-taxable parts of depreciation allowances in firms' EBT outweighs the negative effect due to the non-inclusion of the taxable parts of depreciation allowances in firms' pre-tax cashflow.

3.2.3 Industry Affiliation

I further examine whether or not industry affiliation can be a reliable predictor of firms' true TLCF status. This is because empirical tax research suggests that TLCF tend to be concentrated among certain industries. Auerbach and Poterba (1987), for example, examine a panel of listed U.S. firms between 1981 and 1984. They find that TLCF tend

to be concentrated among firms operating in the industries Oil, Steel, Autos and Airlines. Cooper and Knittel (2010) examine U.S. firms' TLCF status between 1993 and 1999. They conclude that TLCF tend to be concentrated among industries Manufacturing of Durable Goods, Wholesale/Retail, Information Technology and Financial Services.

In order to capture industry affiliation, I rely on firms' four-digit GICS code. Contrary to the other methods I examine, industry affiliation cannot directly be used as a proxy for firms' TLCF status. I compute the method's TLCF status estimates by performing a double-hurdle regression analysis based on a probit model in the first and a truncated normal model in the second tier. In this setting, I determine first whether or not firms are exposed to TLCF, given their industry affiliation. In a second step, I predict the amount of TLCF available for firms classified as having TLCF.

In the probit model, I regress $TAVB_{i,t}$ on a set of $k - 1$ industry dummies:

$$TAVB_{i,t} = \alpha + \sum_{j=1}^{k-1} \beta_j INDUSTRY_{i,j} + \epsilon_{i,t} \quad (2)$$

where $INDUSTRY_{i,j}$ is equal to one if firm i operates in industry j and zero otherwise. Industry k (Information Technology, four-digit GICS code 4500) serves as the reference category. Industries with less than 20 firms in my sample are summarized in a residual category called 'Other'. $\epsilon_{i,t}$ represents the error term. Estimating Equation 2 allows me to predict firm i 's probability of having TLCF available for deduction in year t , given its industry affiliation: $INDPR_{i,t}$. A mathematical derivation of $INDPR_{i,t}$ is shown in Supplemental Appendix A.2. For my further analysis, I set $PAVB_{i,t}^{IND} = 1$ if $INDPR_{i,t} > 0.5$ and hence, the availability of TLCF is more likely than the non-availability of TLCF. If $INDPR_{i,t} \leq 0.5$, $PAVB_{i,t}^{IND} = 0$. The results of the probit regression analysis as well as specifications of $INDPR_{i,t}$ for each of the industry types employed are shown in Supplemental Appendix A.3.

In the truncated normal model, I regress $TAMTD_{i,t}$ on the same set of industry dummies used in Equation 2:

$$TAMTD_{i,t} = \alpha + \sum_{j=1}^{k-1} \beta_j INDUSTRY_{i,j} + \epsilon_{i,t} \quad (3)$$

For a firm-year observation to be included in the truncated normal regression analysis, $TAVB_{i,t}$ has to be equal to one. Estimating Equation 3 thus allows me to predict the amount of TLCF available to TLCF firms, given their industry affiliation: $INDAMTD_{i,t}$.

A mathematical derivation of $INDAMTD_{i,t}$ is shown in Supplemental Appendix A.2. For my further analysis, I set $PAMTD_{i,t}^{IND}$ equal to $INDAMTD_{i,t}$ if $INDPR_{i,t} > 0.5$. If $INDPR_{i,t} \leq 0.5$, $PAMTD_{i,t}^{IND} = 0$. The results of the truncated normal regression as well as specifications of $INDAMTD_{i,t}$ for each of the industry types employed are shown in Supplemental Appendix A.3.

Industry affiliation does not capture any individual firm characteristics. Hence, I expect industry affiliation to perform worse in identifying firms' TLCF status than ILCF or CLCF. In order to see whether or not the method's accuracy can be improved by considering firm-specific characteristics, I also examine the accuracy of industry affiliation in combination with firm age. Relying on firm age as a means of capturing individual firm characteristics seems to be reasonable, because firm age is likely to be an indicator of firms' TLCF status. This is shown by Cooper and Knittel (2010), who point out that TLCF tend to be concentrated among younger firms. I incorporate firm age into my analysis by adding the explanatory variable $AGE_{i,t}$ to both the probit and the truncated normal model shown in Equations 2 and 3, respectively. $AGE_{i,t}$ is defined as the difference between year t and firm i 's incorporation date. The results of the probit and truncated normal regression analyses as well as specifications for $INDPR_{i,t}$ and $INDAMTD_{i,t}$ are shown in Supplemental Appendix A.4. Due to considering firm-specific characteristics, I expect the combination of industry affiliation and firm age to perform better in identifying firms' TLCF status than industry affiliation alone.

3.3 Accuracy Assessment

I assess the accuracy of database-driven methods in classifying firms as either non-TLCF or TLCF firms by determining each method's percentage of correct predictions made: PCP^m . A correct TLCF status prediction is made if $PAVB_{i,t}^m = TAVB_{i,t}$. In order to assess which method is best in predicting the availability of TLCF, I assign a rank based on PCP^m to each method m . I assign the lowest rank to the method with the highest value of PCP^m , and the highest rank to the method with lowest value of PCP^m . Thus, the method with the lowest rank is assumed to be best in predicting whether or not a firm is exposed to TLCF.

For firm-year observations correctly classified as having TLCF (with $TAMTD_{i,t} > 0$ and $PAMTD_{i,t}^m > 0$), I determine logarithmic prediction errors in order to assess the accuracy of database-driven methods in predicting the amount of TLCF available. For method m ,

the logarithmic prediction error made for firm i in year t is defined as follows:

$$LPE_{i,t}^m = \ln(PAMTD_{i,t}^m) - \ln(TAMTD_{i,t}) \quad (4)$$

Due to the fact that I determine $LPE_{i,t}^m$ only for firm-year observations correctly classified as having TLCF, I do not experience any problems with the fact that $\ln(x)$ is defined only for strictly positive values of x . I decide upon which method is most accurate in predicting firms' TLCF amount, by determining the proportion of logarithmic prediction errors that are small in magnitude, $PSLPE^m$, for each method m . For my analysis, I choose an arbitrary threshold level that is relatively close to the optimum error level of zero in order to decide upon whether or not $|LPE_{i,t}^m|$ is small. In particular, I assume that $|LPE_{i,t}^m|$ is small if $|LPE_{i,t}| \leq 0.1$. In such a case, $TAMTD_{i,t}$ is either not more than 1.11 times or not more than 0.9 times as large as $PAMTD_{i,t}^m$. Each method m receives a rank based on its $PSLPE^m$. I assign the lowest rank to the method with the highest value of $PSLPE^m$, and the highest rank to the method with the lowest value of $PSLPE^m$. Thus, the method with the lowest rank is assumed to be most accurate in predicting firms' TLCF amount.

4 Data

4.1 Sample Selection and Overview

I examine the accuracy of database-driven methods in predicting firms' TLCF status based on a panel of listed Italian parent companies between 2010 and 2012. In particular, I consider firms listed in one, two or all of the three observation years. Information on whether or not an Italian firm is listed in a given year is obtained from the Amadeus database. Insights into the sample selection process are provided in Table 1.

The full sample of listed Italian parent companies corresponds to 837 firm-year observations between 2010 and 2012. For my analysis, I do not consider 177 firm-year observations that belong to the financial industry (two-digit GICS code 40). The preliminary sample size is thus equal to 660 firm-year observations. For an observation to be included in my final sample, all data necessary to determine firms' true and each method's predicted TLCF status has to be available. This requirement results in the removal of 335 further firm-year observations, reducing the final sample size to 325 observations from 137 firms. I am not able to derive firms' true TLCF status if no unconsolidated IFRS statement is available (66 firm-year observations), or if $DTA_{i,t-1}$ is disclosed only in combination

Table 1: Sample Selection Process

This table provides insights into the sample selection process.

	N
full sample of listed Italian parent companies	837
– firms operating in the financial industry (40)	– 177
preliminary sample size	660
– unconsolidated IFRS statement is unavailable	– 66
– firms’ TLCF status cannot be reliably determined	– 64
– corporate income tax rate used to determine $DTA_{i,t}$ is not disclosed by firms operating in the industries Energy (1010) or Utilities (5510)	– 10
– missing data in the Amadeus database	– 195
final sample size	325

with other temporary differences (64 firm-year observations). Moreover, as pointed out in Section 3.1, I am not able to determine the TLCF status of firms that belong to the industries Energy or Utilities, and that do not disclose the corporate income tax rate used to determine $DTA_{i,t-1}$ (ten firm-year observations). A method’s predicted TLCF status cannot be derived if the Amadeus database reports a missing value for at least one of the data items required to determine the method’s TLCF status prediction (195 firm-year observations).

186 out of 325 firm-year observations in my sample (57.23%) do not have any TLCF, whereas 139 firm-year observations (42.77%) have. For firm-year observations with TLCF, $TAMTD_{i,t}$ varies between 34 and 162,049 TEUR. The amount of TLCF available to firms can thus be large. In my sample, however, only a small number of firms is exposed to large TLCF. This is shown by the level of the 75%-percentile of $TAMTD_{i,t}$, which corresponds to 31,367 TEUR for TLCF firms. On average, the amount of TLCF available to TLCF firms is equal to 23,184 TEUR. Detailed descriptive statistics regarding firms’ true TLCF status are shown in Supplemental Appendix A.5.

4.2 Reliability of IAS 12.81 Information on TLCF

Table 2 provides information on the extent to which the firms in my sample disclose information on their TLCF status in the notes to their unconsolidated IFRS statement. As shown in Column 1 of Table 2, 82 out of 325 firm-year observations (25.23%) disclose the total stock of TLCF. 54 of these observations report a positive amount of TLCF, whereas 28 observations explicitly state that they are not exposed to any TLCF. This is shown in Columns 2 and 3 of Table 2. Firms that report the total stock of TLCF also disclose all relevant IAS 12.81 information on TLCF. 22 out of 325 observations (6.77%) do not

disclose their TDCF stock, but publish either all TDCF information required by IAS 12.81, or explicitly state that parts of the standard’s requirements are not relevant for them. All of these firms have TDCF. 221 out of 325 firm-year observations (68%) do not disclose all information on TDCF required by IAS 12.81, and do also not explicitly indicate that parts of the standard’s requirements are not relevant for them. 158 of these 221 firm-year observations do not report anything on TDCF. For these firms, I assume that no TDCF are available. Overall, 104 out of 325 firm-year observations (32%) offer reliable IFRS statement information on firms’ TDCF status, whereas 221 firm-year observations (68%) do not. This finding supports previous conclusions drawn in empirical tax research, which suggest that many firms do not disclose all information on TDCF required by IAS 12.81 (Kager and Niemann, 2013, Petermann and Schanz, 2013). As pointed out in Section 3.1, I assume that the latter firms provide all TDCF information relevant for them. The results of the survey performed in order to verify the accuracy of this assumption are discussed in detail below.

Table 2: IFRS Statement Information on TDCF

This table provides insights into the extent to which firms disclose information on their TDCF status in the notes to their unconsolidated IFRS statement. $TAVB_{i,t}$ is equal to one if firm i has TDCF for deduction in year t and zero otherwise.

	(1) Total		(2) $TAVB_{i,t} = 1$		(3) $TAVB_{i,t} = 0$	
	N	%	N	%	N	%
Disclosure of:						
TDCF stock	82	25.23%	54	35.85%	28	15.05%
full IAS 12.81 information on TDCF	22	6.77%	22	15.83%	0	0.00%
partial/no IAS 12.81 information on TDCF	221	68.00%	63	45.32%	158	84.95%
Total	325	100.00%	139	42.77%	186	57.23%

Only seven out of 137 firms (fourteen firm-year observations) took part in the survey. The findings derived below can thus only be a weak indicator of the accuracy of my assumption regarding firms’ reporting behavior. For all fourteen firm-year observations, $TAVB_{i,t}$, as determined based on IFRS statement information, corresponds perfectly to $TAVB_{i,t}$, as revealed in the survey. Predicting the availability of TDCF based on IAS 12.81 information thus seems to be accurate. For firms with TDCF, $TAMTD_{i,t}$, as determined based on IFRS statement information, and $TAMTD_{i,t}$, as revealed in the survey, are compared in Table 3.

Table 3: Amount of TLCF - Survey Results

For TLCF firms, this table compares $TAMTD_{i,t}$, as determined based on IFRS statement information (Column 1), to $TAMTD_{i,t}$, as revealed in the survey regarding firms' true TLCF status (Column 2). $TAMTD_{i,t}$ represents firm i 's amount of TLCF available for deduction in year t in TEUR. The logarithmic prediction error shown in Column 3 corresponds to the difference between the logarithm of $TAMTD_{i,t}$ shown in Column 1 and the logarithm of $TAMTD_{i,t}$ shown in Column 2.

(1) $TAMTD_{i,t}$ IFRS Statement	(2) $TAMTD_{i,t}$ Survey	(3) Logarithmic Prediction Error
6,000	11,360	-0.638
22,480	30,767	-0.314
29,724	34,748	-0.156
32,015	37,078	-0.147
14,538	16,808	-0.145
45,349	46,338	-0.022
23,622	23,622	0.000
12,001	11,886	0.010
7,993	7,868	0.016
100,418	95,665	0.048
49,886	41,182	0.192

In order to assess the accuracy of my assumption regarding firms' reporting behavior, I determine a logarithmic prediction error, which is shown in Column 3 of Table 3. The error corresponds to the difference between the logarithm of $TAMTD_{i,t}$, as determined based on IFRS statement information (Column 1), and the logarithm of $TAMTD_{i,t}$, as revealed in the survey (Column 2). For one firm-year observation, $TAMTD_{i,t}$ based on IFRS statement information corresponds perfectly to the amount revealed. For six observations, $TAMTD_{i,t}$, as determined based on IFRS statement information, is below $TAMTD_{i,t}$, as revealed in the survey. This indicates that some firms do indeed not fully follow the provisions imposed by IAS 12.81. For four firm-year observations, $TAMTD_{i,t}$ based on IFRS statement information is above the TLCF amount revealed. This unexpected finding might be due to TLCF corrections made after the publishing date. Overall, the deviations from the amount revealed tend to be small. With the exception of the firm-year observations shown in the first two rows of Table 3, the logarithmic prediction error varies between -0.156 and 0.192 . In the former (latter) case, $TAMTD_{i,t}$ based on IFRS statement information is about 0.86 (1.21) times as high as the amount revealed. This indicates that predicting the amount of TLCF available to firms by relying on IAS 12.81 information tends to be quite accurate too. For my analysis, I employ $TAMTD_{i,t}$, as revealed in the survey, instead of $TAMTD_{i,t}$ based on IFRS statement information if both figures differ.

5 Results

5.1 Availability of TLCF

Table 4 shows the accuracy of the methods examined in predicting the availability of TLCF.

Table 4: Availability of TLCF - Results

This table shows the results derived regarding the accuracy of database-driven methods in predicting the availability of TLCF at single-firm level. The findings are based on a total of 325 firm-year observations of listed Italian parent companies between 2010 and 2012. PCP^m is defined as the percentage of correct TLCF status predictions made by method m . The methods are ranked based on the level of PCP^m . The lowest (highest) rank is assigned to the method with the highest (lowest) level of PCP^m .

Method m	(1) PCP^m	(2) Rank m
$ILCF^{EBT,n=1}$	73.85%	5
$ILCF^{EBT,n=2}$	76.31%	3
$ILCF^{EBT,n=3}$	78.15%	2
$ILCF^{EBT,n=4}$	79.08%	1
$ILCF^{EBIT,n=1}$	71.69%	11
$ILCF^{EBIT,n=2}$	72.62%	8
$ILCF^{EBIT,n=3}$	72.62%	8
$ILCF^{EBIT,n=4}$	73.85%	5
$ILCF^{EBITDA,n=1}$	68.62%	16
$ILCF^{EBITDA,n=2}$	69.85%	15
$ILCF^{EBITDA,n=3}$	70.15%	13
$ILCF^{EBITDA,n=4}$	70.15%	13
$CLCF^{n=1}$	71.08%	12
$CLCF^{n=2}$	72.31%	10
$CLCF^{n=3}$	73.85%	5
$CLCF^{n=4}$	74.77%	4
<i>Industry</i>	60.92%	17
<i>Industry/Age</i>	59.38%	18

PCP^m , as shown in Column 1 of Table 4, is highest for ILCF based on firms' EBT from four past years (79.08%). It is lowest for the method based on industry affiliation and firm age (59.38%). For all other methods, PCP^m varies between 59.38% and 79.08%. Table 5 shows how the accuracy of ILCF and CLCF changes if the degree of past earnings/cashflow information considered is increased. It reports the change in PCP^m , $\Delta PCP_{n=1}^m$, if method m is based on information from n (with $2 \leq n \leq 4$) instead of only one past year. My judgments regarding the impact of an increase in the degree of past earnings/cashflow information considered are based on the results of a chi²-test of independency.

As shown in Table 5, the accuracy of ILCF and CLCF does not significantly change if the degree of past earnings/cashflow information considered is increased. Hence, it

Table 5: Availability of TLCF - Increase in the Degree of Past Information Considered

This table shows how the accuracy of ILCF and CLCF changes if the degree of past earnings/cashflow information considered is increased. $\Delta PCP_{n=1}^m$ (in percentage points) corresponds to the change in PCP^m if method m is based on information from n (with $2 \leq n \leq 4$) instead of only one past year. The values used to determine $\Delta PCP_{n=1}^m$ as well as a definition of PCP^m can be found in Table 4. Judgments regarding the impact of an increase in the degree of past earnings/cashflow information considered are based on the results of a chi²-test of independency. *, ** and *** correspond to significance levels of 10%, 5% and 1%, respectively.

$\Delta PCP_{n=1}^m$	(1) $n = 2$	(2) $n = 3$	(3) $n = 4$
$ILCF^{EBT}$	+2.46	+4.31	+5.23
$ILCF^{EBIT}$	+0.92	+0.92	+2.15
$ILCF^{EBITDA}$	+1.23	+1.54	+1.54
$CLCF$	+1.23	+2.77	+3.69

seems not to matter whether ILCF or CLCF are based on information from only one or a higher number of years. This finding indicates that the positive accuracy effect due to the inclusion of more relevant past information tends to be about as large as the negative accuracy effect, which results from the induced increase in the level of book-tax differences (see Section 3.2). Table 6 shows whether or not ILCF based on firms' EBIT or EBITDA or CLCF are more accurate in predicting the availability of TLCF than ILCF based on firms' EBT. It reports the change in PCP^m , ΔPCP_{EBT}^m , if method m is based on firms' EBIT, EBITDA or pre-tax cashflow instead of firms' EBT from n past years (with $1 \leq n \leq 4$). My judgments regarding the ability of ILCF based on firms' EBIT or EBITDA or CLCF to outperform ILCF based on firms' EBT are based on the results of a chi²-test of independency.

As shown in Table 6, I do not find a significant difference in the accuracy of ILCF based on firms' EBT and ILCF based on firms' EBIT. Hence, it seems not to matter whether ILCF are based on firms' EBT or EBIT. The same applies to CLCF. I do find a significant difference in the accuracy of ILCF based on firms' EBT and ILCF based on firms' EBITDA if $n \geq 2$. As shown in Column 3 of Table 6, the latter perform significantly worse in predicting the availability of TLCF than the former. ΔPCP_{EBT}^m varies between -6.46 and -8.92 percentage points. ILCF based on firms' EBT should therefore be preferred over ILCF based on firms' EBITDA. This finding indicates that the negative accuracy effect due to the inclusion of financial income and depreciation allowances in firms' EBT tends to be significantly smaller than the negative accuracy effect due to the non-inclusion of both in firms' EBITDA (see Section 3.2). A comparison between ILCF and CLCF and the methods based on industry affiliation is made in Table 7. In Columns 1-4, Table 7 reports the change in PCP^m , ΔPCP_{IND}^m , if ILCF or CLCF and not the method based on

industry affiliation (Panel A) or industry affiliation and firm age (Panel B) are employed. In Column 5, Table 7 also reports how the inclusion of firm age affects the performance of industry affiliation in predicting the availability of TLCF. In this case, ΔPCP_{IND}^m corresponds to the difference between PCP^m of the method based on industry affiliation and firm age and PCP^m of the method based on industry affiliation (and vice versa). My judgments regarding the differences in the methods' accuracy are based on the results of a chi²-test of independency.

Table 6: Availability of TLCF - Different Earnings/Cashflow Measures

This table shows whether or not ILCF based on firms' EBIT or EBITDA or CLCF are more accurate in predicting the availability of TLCF than ILCF based on firms' EBT. ΔPCP_{EBT}^m (in percentage points) corresponds to the change in PCP^m if method m is based on firms' EBIT, EBITDA or pre-tax cashflow instead of firms' EBT from n past years (with $1 \leq n \leq 4$). The values used to determine ΔPCP_{EBT}^m as well as a definition of PCP^m can be found in Table 4. Judgments regarding the ability of ILCF based on firms' EBIT or EBITDA or CLCF to outperform ILCF based on firms' EBT are based on the results of a chi²-test of independency. *, ** and *** correspond to significance levels of 10%, 5% and 1%, respectively.

ΔPCP_{EBT}^m	(1) $ILCF^{EBIT}$	(2) $ILCF^{EBITDA}$	(3) $CLCF$
$n = 1$	-1.23	-5.23	-2.77
$n = 2$	-3.69	-6.46 *	-4.00
$n = 3$	-5.54	-8.00 **	-4.31
$n = 4$	-5.23	-8.92 ***	-4.31

As shown in Columns 1-4 of Table 7, the methods based on industry affiliation perform significantly worse than ILCF or CLCF. For ILCF based on firms' EBT for example, ΔPCP_{IND}^m varies between +12.91 and +18.15 percentage points (Panel A) and between +14.46 and +19.69 percentage points (Panel B). This finding suggests that there is hardly any concentration of firms with or without TLCF among certain types of industry. Accordingly, I do not recommend the use of industry affiliation in order to predict the availability of TLCF. The inclusion of firm age does not help to improve the method's accuracy. This is shown in Column 5 of Table 7. Firm age thus seems not be a reliable predictor of firms' TLCF status.

In order to see whether or not certain firm characteristics and/or the year of observation influence the probability of making a wrong TLCF status prediction, I perform the following regression analysis based on a probit model for each method m :

$$PE_{i,t}^m = \alpha + \beta_1 \ln(SIZE_{i,t-1}) + \beta_2 AGE_{i,t} + \sum_{j=1}^{k-1} \gamma_j INDUSTRY_{i,j} + \delta_1 Y11_t + \delta_2 Y12_t + \epsilon_{i,t} \quad (5)$$

Table 7: Availability of TLCF - Industry Affiliation

This table shows whether or not the methods based on industry affiliation are more accurate in predicting the availability of TFCF than ILCF or CLCF (Columns 1-4). ΔPCP_{IND}^m (in percentage points) corresponds to the difference in PCP^m if ILCF or CLCF and not the method based on industry affiliation (Panel A) or industry affiliation and firm age (Panel B) are employed. This table also shows whether or not the accuracy of industry affiliation can be enhanced by considering firm age (Column 5). In this case, ΔPCP_{IND}^m corresponds to the difference between PCP^m of the method based on industry affiliation and firm age and PCP^m of the method m based on industry affiliation (and vice versa). The values used to determine ΔPCP_{IND}^m as well as a definition of PCP^m can be found in Table 4. Judgments regarding the differences in the methods' accuracy are based on the results of a chi²-test of independency. *, ** and *** correspond to significance levels of 10%, 5% and 1%, respectively.

(A) Industry					
ΔPCP_{IND}^m	(1) $ILCF^{EBT}$	(2) $ILCF^{EBIT}$	(3) $ILCF^{EBITDA}$	(4) $CLCF$	(5) $Industry/Age$
$n = 1$	+12.92 ***	+10.77 ***	+7.69 **	+10.15 ***	-1.54
$n = 2$	+15.38 ***	+11.69 ***	+8.92 **	+11.38 ***	
$n = 3$	+17.23 ***	+11.69 ***	+9.23 **	+12.92 ***	
$n = 4$	+18.15 ***	+12.92 ***	+9.23 **	+13.85 ***	
(B) Industry and Firm Age					
ΔPCP_{IND}^m	(1) $ILCF^{EBT}$	(2) $ILCF^{EBIT}$	(3) $ILCF^{EBITDA}$	(4) $CLCF$	(5) $Industry$
$n = 1$	+14.46 ***	+12.31 ***	+9.23 **	+10.15 ***	+1.54
$n = 2$	+16.92 ***	+13.23 ***	+10.46 **	+11.38 ***	
$n = 3$	+18.77 ***	+13.23 ***	+10.77 **	+12.92 ***	
$n = 4$	+19.69 ***	+14.46 ***	+10.77 **	+13.85 ***	

where $PE_{i,t}^m$ is a dummy variable, which is equal to one if method m wrongly predicts firm i 's TFCF status in year t and zero otherwise. $SIZE_{i,t-1}$ represents firm size. It is defined as firm i 's amount of total assets at the end of year $t - 1$. $AGE_{i,t}$ and $INDUSTRY_{i,j}$ are specified as in Section 3.2.3. In order to capture a potential impact of the observation year, I employ two year dummies. $Y11_t$ ($Y12_t$) is equal to one in year 2011 (2012) and zero otherwise. Year 2010 serves as the reference category. For the method that is most accurate in predicting the availability of TFCF, the ILCF based on firms' EBT from four past years, the results of the probit regression analysis are shown in Table 8. For all other methods, regression results are provided in Supplemental Appendix A.6. Table 8 reports both coefficient estimates and the average partial effects (APE) of the explanatory variables on the probability of making a TFCF status prediction error.

As shown in Table 8, the probability of making a wrong TFCF status prediction is significantly higher for firms operating in the industries Capital Goods, Consumer Durables & Apparel and Media than for firms operating in the industry type Information Technology. Thus, the ILCF of the former deviate more strongly from firms' TFCF than the ILCF of the latter. The magnitude of the impact of industry affiliation on the probability of

making a wrong TLCF status prediction tends to be substantial. For firms operating in the industry Capital Goods, the probability of making a wrong TLCF status prediction is on average +19.3 percentage points higher than for firms operating in the industry Information Technology. For industry types Consumer Durables & Apparel and Media, the average difference in the probability of making a prediction error is equal to +20.5 and +23.2 percentage points, respectively. This suggests that the accuracy of ILCF based on firms' EBT from four past years suffers substantially in settings where a large number of firms operating in the industries Capital Goods, Consumer Durables & Apparel or Media is employed. In such settings, I thus strongly recommend not to rely on ILCF based on firms' EBT from four past years as a means of predicting firms' TLCF status. There are no significant effects of firm size, firm age or the year of observation on the probability of making a wrong TLCF status prediction.

Overall, the results shown in Table 4 suggest that database-driven methods do not perform particularly well in predicting the availability of TLCF. Even for the most accurate of the methods examined, more than 20% of the TLCF status predictions tend to be wrong. This fraction is likely to increase even further in settings where a large number of firms operating in the industries Capital Goods, Consumer Durables & Apparel or Media is employed. The latter finding has important implications for empirical tax research. It suggests that studies relying on database-driven methods similar to the ones tested here might not be able to correctly classify firms as either TLCF or non-TLCF firms. This reduces the studies' ability to detect any differences in the behavior of both types of firms. Any results obtained regarding the impact of TLCF are thus likely to be distorted. In line with these arguments, I cannot recommend the use of database-driven identification methods in empirical tax research.

Previous empirical studies that rely on database-driven methods similar to the ones tested here include the investigations of Bernasconi et al. (2005), Overesch and Voeller (2010), Overesch and Wamser (2010), Buettner et al. (2011), Buettner et al. (2012), Dreßler and Overesch (2013), Stöckl and Winner (2013) and Overesch and Wamser (2014). In order to verify the accuracy of the conclusions drawn in these studies, I recommend conducting replication studies using firms' true TLCF status instead of database-driven proxies. For future studies, I recommend not to rely on database-driven identification methods at all. In empirical tax research, there is thus a strong need for firm-specific TLCF status information provided by local tax authorities.

Table 8: Availability of TLCF - Error Analysis for ILCF based on Firms' EBT from Four Past Years

This table shows how firm characteristics and the observation year influence the probability of making a wrong TLCF status prediction for ILCF based on firms' EBT from four past years. The results are based on the following probit regression model: $PE_{i,t}^m =$

$$\alpha + \beta_1 \ln(SIZE_{i,t-1}) + \beta_2 AGE_{i,t} + \sum_{j=1}^{k-1} \gamma_j INDUSTRY_{i,j} + \delta_1 Y11_t + \delta_2 Y12_t + \epsilon_{i,t}.$$

$PE_{i,t}^m$ is equal to one if method m wrongly predicts firm i 's TLCF status in year t and zero otherwise. $SIZE_{i,t-1}$ is defined as the amount of total assets at the end of year $t-1$. $AGE_{i,t}$ corresponds to the difference between year t and firm i 's incorporation year. $INDUSTRY_{i,j}$ is equal to one if firm i operates in industry j and zero otherwise. Industry type Other summarizes industries Energy (1010) and Telecommunication Services (5010). Industry Information Technology (4500) serves as the reference category. $Y11_t$ ($Y12_t$) is equal to one in year 2011 (2012) and zero otherwise. Year 2010 serves as the reference category. $\epsilon_{i,t}$ represents the error term. Standard errors are shown in parentheses. They are clustered at firm-level. *, ** and *** correspond to significance levels of 10%, 5% and 1%, respectively.

	(1) Coefficient	(2) APE
$\ln(SIZE_{i,t-1})$	0.079 (0.059)	0.021
$AGE_{i,t}$	0.002 (0.004)	0.000
<i>Other (0)</i>	0.130 (0.437)	0.035
<i>Materials (1510)</i>	0.792 (0.515)	0.213
<i>Industrials Other (2000)</i>	0.461 (0.595)	0.124
<i>Capital Goods (2010)</i>	0.720 * (0.382)	0.193
<i>Consumer Discretionary Other (2500)</i>	0.185 (0.518)	0.050
<i>Consumer Durables & Apparel (2520)</i>	0.763 * (0.421)	0.205
<i>Media (2540)</i>	0.863 * (0.453)	0.232
<i>Consumer Staples (3000)</i>	0.475 (0.465)	0.128
<i>Health Care (3500)</i>	0.013 (0.468)	0.004
<i>Utilities (5510)</i>	-0.087 (0.445)	-0.023
$Y11_t$	-0.150 (0.157)	-0.040
$Y12_t$	-0.144 (0.168)	-0.039
<i>Constant</i>	-2.292 *** (0.784)	
N	325	
Wald statistic	16.95	
Pseudo R ²	0.065	

5.2 Amount of TLCF

Table 9 shows the accuracy of the methods examined in predicting the amount of TFCF available to firms. For each method m , the results discussed below are based on those firm-year observations that are correctly classified as having TFCF. My results thus critically depend upon the findings derived in Section 5.1.

Table 9: Amount of TFCF - Results

This table shows the results derived regarding the accuracy of database-driven methods in predicting the amount of TFCF available at single-firm level. For each method m , the results shown below are based on those firm-year observations that are correctly classified as having TFCF. They thus critically depend upon the results derived in Section 5.1. $PSLPE^m$ is defined as method m 's proportion of logarithmic prediction errors that are small in magnitude. The methods are ranked based on the level of $PSLPE^m$. The lowest (highest) rank is assigned to the method with the highest (lowest) level of $PSLPE^m$.

Method m	(1) N	(2) $PSLPE^m$	(3) Rank m
$ILCF^{EBT,n=1}$	96	3.13%	17
$ILCF^{EBT,n=2}$	107	6.54%	12
$ILCF^{EBT,n=3}$	113	7.96%	10
$ILCF^{EBT,n=4}$	116	7.76%	11
$ILCF^{EBIT,n=1}$	102	5.88%	13
$ILCF^{EBIT,n=2}$	109	10.09%	8
$ILCF^{EBIT,n=3}$	112	16.96%	3
$ILCF^{EBIT,n=4}$	117	15.38%	4
$ILCF^{EBITDA,n=1}$	79	3.80%	15
$ILCF^{EBITDA,n=2}$	85	3.53%	16
$ILCF^{EBITDA,n=3}$	87	10.34%	6
$ILCF^{EBITDA,n=4}$	87	12.64%	5
$CLCF^{n=1}$	75	0.00%	18
$CLCF^{n=2}$	82	4.88%	14
$CLCF^{n=3}$	87	10.34%	6
$CLCF^{n=4}$	90	8.89%	9
<i>Industry</i>	56	19.64%	1
<i>Industry/Age</i>	56	19.64%	1

$PSLPE^m$, as shown in Column 2 of Table 9, is highest for the methods based on industry affiliation (19.64%). It is lowest for CLCF based on pre-tax cashflow from one past year (0%). For all other methods, $PSLPE^m$ varies between 0% and 19.64%. Table 10 shows how the accuracy of ILCF and CLCF changes if the degree of past earnings/cashflow information considered is increased. It reports the change in $PSLPE^m$, $\Delta PSLPE_{n=1}^m$, if method m is based on information from n (with $2 \leq n \leq 4$) instead of only one past year. My judgments regarding the impact of an increase in the degree of past information

considered are based on the results of a chi²-test of independency or, alternatively, Fisher's exact test.

Table 10: Amount of TLCF - Increase in the Degree of Past Information Considered

This table shows how the accuracy of ILCF and CLCF changes if the degree of past earnings/cashflow information considered is increased. $\Delta PSLPE_{n=1}^m$ (in percentage points) corresponds to the change in $PSLPE^m$ if method m is based on information from n (with $2 \leq n \leq 4$) instead of only one past year. The values used to determine $\Delta PSLPE_{n=1}^m$ as well as a definition of $PSLPE^m$ can be found in Table 9. Judgments regarding the impact of an increase in the degree of past earnings/cashflow information considered are based on the results of a chi²-test of independency or, alternatively, Fisher's exact test (^a). *, ** and *** correspond to significance levels of 10%, 5% and 1%, respectively.

$\Delta PSLPE_{n=1}^m$	(1) $n = 2$	(2) $n = 3$	(3) $n = 4$
$ILCF^{EBT}$	+3.42 ^a	+4.84	+4.63
$ILCF^{EBIT}$	+4.21	+11.08 **	+9.50 **
$ILCF^{EBITDA}$	-0.27 ^a	+6.55	+8.85 **
$CLCF$	+4.88	+10.34 ***	+8.89 *** ^a

As shown in Table 10, the accuracy of ILCF based on firms' EBT does not significantly change if the degree of past earnings information considered is increased. Hence, it seems not to matter whether ILCF based on firms' EBT are determined using information from only one or a higher number of years. In contrast to that, the accuracy of ILCF based on firms' EBIT and CLCF significantly increases if the methods are based on earnings/cashflow information from three or four instead of only one past year. For ILCF based on firms' EBIT, $\Delta PSLPE_{n=1}^m$ varies between +9.50 and +11.08 percentage points. For CLCF, $\Delta PSLPE_{n=1}^m$ varies between +8.89 and +10.34 percentage points. For ILCF based on firms' EBITDA, a significant increase in the method's accuracy can be achieved by relying on earnings information from four instead of only one past year. The increase is equal to +8.85 percentage points. ILCF based on firms' EBIT or EBITDA and CLCF should therefore be based on information from at least three or four years in the past. For these methods, the positive accuracy effect due to the inclusion of more relevant past earnings/cashflow information tends to be significantly larger than the negative accuracy effect, which results from the induced increase in the level of book-tax differences. Table 11 shows whether or not ILCF based on firms' EBIT or EBITDA or CLCF are more accurate in predicting the amount of TLCF than ILCF based on firms' EBT. It reports the change in $PSLPE^m$, $\Delta PSLPE_{EBT}^m$, if method m is based on firms' EBIT, EBITDA or pre-tax cashflow instead of firms' EBT from n past years (with $1 \leq n \leq 4$). My judgments regarding the ability of ILCF based on firms' EBIT or EBITDA or CLCF to outperform ILCF based on firms' EBT are based on the results of a chi²-test of independency or, alternatively, Fisher's exact test.

Table 11: Amount of TLCF - Different Earnings/Cashflow Measures

This table shows whether or not ILCF based on firms' EBIT or EBITDA or CLCF are more accurate in predicting the amount of TLCF than ILCF based on firms' EBT. $\Delta PSLPE_{EBT}^m$ (in percentage points) corresponds to the change in $PSLPE^m$ if method m is based on firms' EBIT, EBITDA or pre-tax cashflow instead of firms' EBT from n past years (with $1 \leq n \leq 4$). The values used to determine $\Delta PSLPE_{EBT}^m$ as well as a definition of $PSLPE^m$ can be found in Table 9. Judgments regarding the ability of ILCF based on firms' EBIT or EBITDA or CLCF to outperform ILCF based on firms' EBT are based on the results of a chi²-test of independency or, alternatively, Fisher's exact test (^a). *, ** and *** correspond to significance levels of 10%, 5% and 1%, respectively.

$\Delta PSLPE_{EBT}^m$	(1) $ILCF^{EBIT}$	(2) $ILCF^{EBITDA}$	(3) $CLCF$
$n = 1$	+2.76 ^a	+0.67 ^a	-3.13 ^a
$n = 2$	+3.55	-3.01 ^a	-1.66 ^a
$n = 3$	+9.00 **	+2.38	+2.38
$n = 4$	+7.63 **	+4.89	+1.13

As shown in Table 11, I do not find a significant difference in the accuracy of ILCF based on firms' EBT and ILCF based on firms' EBITDA. Hence, it seems not to matter whether ILCF are based on firms' EBT or EBITDA. The same applies to CLCF. I do find a significant difference in the accuracy of ILCF based on firms' EBT and ILCF based on firms' EBIT if $n \geq 3$. As shown in Column 1 of Table 11, the former perform significantly worse in predicting the amount of TLCF available than the latter. $\Delta PSLPE_{EBT}^m$ varies between +7.63 and +9.00 percentage points. This finding indicates that the negative accuracy effect due to the inclusion of financial income in firms' EBT tends to be significantly larger than the negative accuracy effect due to the non-inclusion of financial income in firms' EBIT. It suggests that ILCF based on firms' EBIT should be preferred over ILCF based on firms' EBT in predicting the amount of TLCF available to firms. The latter conclusion is supported by the fact that there is no significant difference in the performances of ILCF based on firms' EBT and ILCF based on firms' EBIT in predicting the availability of TLCF (see Section 5.1). A comparison between ILCF and CLCF and the methods based on industry affiliation is made in Table 12. In Columns 1-4, Table 7 reports the change in $PSLPE^m$, $\Delta PSLPE_{IND}^m$, if ILCF or CLCF and not the method based on industry affiliation (Panel A) or industry affiliation and firm age (Panel B) are employed. In Column 5, Table 12 also reports how the inclusion of firm age affects the performance of industry affiliation in predicting the amount of TLCF available. In this case, $\Delta PSLPE_{IND}^m$ corresponds to the difference between $PSLPE^m$ of the method based on industry affiliation and firm age and $PSLPE^m$ of the method based on industry affiliation (and vice versa). My judgments regarding the differences in the methods' accuracy are based on the results of a chi²-test of independency.

Table 12: Amount of TLCF - Industry Affiliation

This table shows whether or not the methods based on industry affiliation are more accurate in predicting the amount of TLCF than ILCF or CLCF (Columns 1-4). $\Delta PSLPE_{IND}^m$ (in percentage points) corresponds to the difference in $PSLPE^m$ if ILCF or CLCF and not the method based on industry affiliation (Panel A) or industry affiliation and firm age (Panel B) are employed. This table also shows whether or not the accuracy of industry affiliation can be enhanced by considering firm age (Column 5). In this case, $\Delta PSLPE_{IND}^m$ corresponds to the difference between $PSLPE^m$ of the method based on industry affiliation and firm age and $PSLPE^m$ of the method m based on industry affiliation (and vice versa). The values used to determine $\Delta PSLPE_{IND}^m$ as well as a definition of $PSLPE^m$ can be found in Table 9. Judgments regarding the differences in the methods' accuracy are based on the results of a chi²-test of independency. *, ** and *** correspond to significance levels of 10%, 5% and 1%, respectively.

(A) Industry					
$\Delta PSLPE_{IND}^m$	(1) $ILCF^{EBT}$	(2) $ILCF^{EBIT}$	(3) $ILCF^{EBITDA}$	(4) $CLCF$	(5) $Industry/Age$
$n = 1$	-16.52 ***	-13.76 ***	-15.85 ***	-19.64 ***	0.00
$n = 2$	-13.10 **	-9.55 *	-16.11 ***	-14.76 ***	
$n = 3$	-11.68 **	-2.68	-9.30	-9.30	
$n = 4$	-11.88 **	-4.26	-7.00	-10.75 *	

(B) Industry and Firm Age					
$\Delta PSLPE_{IND}^m$	(1) $ILCF^{EBT}$	(2) $ILCF^{EBIT}$	(3) $ILCF^{EBITDA}$	(4) $CLCF$	(5) $Industry$
$n = 1$	-16.52 ***	-13.76 ***	-15.85 ***	-19.64 ***	0.00
$n = 2$	-13.10 **	-9.55 *	-16.11 ***	-14.76 ***	
$n = 3$	-11.68 **	-2.68	-9.30	-9.30	
$n = 4$	-11.88 **	-4.26	-7.00	-10.75 *	

As shown in Columns 1-4 of Table 12, the methods based on industry affiliation perform significantly better than ILCF based on firms' EBT. $\Delta PSLPE_{IND}^m$ varies between -11.68 and -16.52 percentage points (Panels A and B). The former also perform significantly better in predicting firms' TLF amount than ILCF based on firms' EBIT or EBITDA if $n \leq 2$. For ILCF based on firms' EBIT for example, $\Delta PSLPE_{IND}^m$ varies between -9.55 and -13.76 percentage points. CLCF perform significantly worse than the methods based on industry affiliation if $n \neq 3$. $\Delta PSLPE_{IND}^m$ varies between -10.75 and -19.64 percentage points. These findings suggest that within specific types of industry, the amount of TLF available to firms tends to be alike. However, I strongly recommend not to rely on the methods based on industry affiliation as a means of predicting firms' TLF amount. This is due to the fact that these methods perform worst in predicting the availability of TLF (see Section 5.1). As in Section 5.1, the inclusion of firm age does not help to improve the accuracy of the method based on industry affiliation. This is shown in Column 5 of Table 12.

In order to see whether or not certain firm characteristics and/or the year of observation influence the probability of making a logarithmic prediction error that is large in mag-

nitude, I perform regression analyses based on a variant of the probit model specified in Equation 5. For the analysis in this section, the dependent variable of the probit model is replaced by a dummy variable, $LLPE_{i,t}^m$, which is equal to one if $|LPE_{i,t}^m| > 0.1$ and zero otherwise. For the method that is most accurate in predicting the amount of TLCF available, given its performance in Section 5.1, the ILCF based on firms' EBIT from three past years, the results of the probit regression analysis are shown in Table 13. For all other methods, regression results are provided in Supplemental Appendix A.7.

As shown in Table 13, the probability of making a large logarithmic prediction error tends to be significantly smaller for firm-year observations in 2011 than for firm-year observations in 2010. The impact of the observation year on the method's accuracy tends to be substantial. For firm-year observations in 2011, the probability of making a large logarithmic prediction error is on average -17.90 percentage points smaller than for firm-year observations in 2010. The method's accuracy thus seems to vary substantially by year. I do not find any significant effects of size, firm age or industry affiliation.

Overall, the results shown in Table 10 suggest that database-driven methods perform badly in predicting the amount of TLCF available to firms. Even for the most accurate method, given its performance in predicting the availability of TLCF, more than 80% of the TLCF amount predictions tend to deviate sharply from firms' true TLCF amount. Furthermore, the method's accuracy varies substantially by year of observation. This finding has important implications for empirical tax research. It indicates that studies relying on database-driven methods similar to the ones tested here might not be able to properly determine firms' TLCF amount. Any results obtained regarding the impact of TLCF are thus likely to be distorted. For this reason, I recommend not to use database-driven identification methods in empirical tax research.

To my knowledge, there is no study that relies on database-driven methods similar to the ones tested here in order to identify the amount of TLCF available to firms. For example, Mackie-Mason (1990), who examines the impact of TLCF on the probability of making public debt or equity issues, relies on Compustat's data item # 52 in order to determine firms' TLCF amount. Hence, I am not able to draw any conclusions regarding the accuracy of previous empirical findings on TLCF and firm behavior. Nevertheless, I recommend that future studies should not rely on database-driven identification methods at all. In line with the findings shown in Section 5.1, this emphasizes the need for firm-specific TLCF status information provided by local tax authorities in empirical tax research.

Table 13: Amount of TLCF - Error Analysis for ILCF based on Firms' EBIT from Three Past Years

This table shows how firm characteristics and the observation year influence the probability of making a wrong TLCF status prediction for ILCF based on firms' EBIT from three past years. The results are based on the following probit regression model: $LLPE_{i,t}^m = \alpha + \beta_1 \ln(SIZE_{i,t-1}) + \beta_2 AGE_{i,t} + \sum_{j=1}^{k-1} \gamma_j INDUSTRY_{i,j} + \delta_1 Y11_t + \delta_2 Y12_t + \epsilon_{i,t}$. $LLPE_{i,t}^m$ is equal to one if $|LPE_{i,t}^m| > 0.1$ and zero otherwise. $LPE_{i,t}^m$ corresponds to the difference between the logarithm of the TLCF amount predicted by method m and the logarithm of firm i 's true TLCF amount. The definitions of $SIZE_{i,t-1}$, $AGE_{i,t}$, $INDUSTRY_{i,j}$, $Y11_t$, $Y12_t$ and $\epsilon_{i,t}$ are shown in Table 8. Standard errors are shown in parentheses. They are clustered at firm-level. *, ** and *** correspond to significance levels of 10%, 5% and 1%, respectively. ^aOmitted due to perfect prediction.

	(1) Coefficient	(2) APE
$\ln(SIZE_{i,t-1})$	0.024 (0.173)	0.006
$AGE_{i,t}$	-0.010 (0.007)	-0.003
<i>Other (0)</i>	0.158 (0.635)	0.040
<i>Materials (1510)</i>	(om.) ^a	
<i>Industrials Other (2000)</i>	0.428 (0.715)	0.108
<i>Capital Goods (2010)</i>	-0.399 (0.484)	-0.101
<i>Consumer Discretionary Other (2500)</i>	0.643 (0.736)	0.163
<i>Consumer Durables & Apparel (2520)</i>	0.748 (0.652)	0.189
<i>Media (2540)</i>	-0.122 (0.631)	-0.031
<i>Consumer Staples (3000)</i>	(om.) ^a	
<i>Health Care (3500)</i>	0.016 (0.496)	0.004
<i>Utilities (5510)</i>	(om.) ^a	
$Y11_t$	-0.707 * (0.381)	-0.179
$Y12_t$	-0.503 (0.415)	-0.127
<i>Constant</i>	1.152 (2.027)	
N	93	
Wald statistic	12.80	
Pseudo R ²	0.107	

6 Summary and Outlook

In this paper, I provide empirical evidence on the accuracy of database-driven methods in predicting the availability and the amount of TLCF at single-firm level. Despite the fact that the majority of empirical studies on TLCF and firm behavior relies on database-driven identification methods, empirical evidence on their accuracy tends to be scarce. My analysis thus offers insights highly relevant for empirical tax research. The methods I examine are ILCF, CLCF and methods based on industry affiliation. The fact that these methods are solely based on standard information, which is typically available in financial databases, ensures that my results tend to be relevant for studies employing various types of databases. The fact that I investigate the accuracy of ILCF further implies that my results tend to be highly relevant for studies examining the impact of TLCF in countries that either prescribe or allow the appliance of the IFRS, given that their institutional setting is similar to the one employed here. The same applies to countries with accounting standards similar to the IFRS.

My investigation is based on a panel of listed Italian parent companies between 2010 to 2012 (325 firm-year observations). In order to assess the methods' accuracy, I compare firms' true TLCF status, as determined based on IFRS statement information, to the TLCF status predictions of the methods examined. In order to make judgments regarding the accuracy of database-driven methods in predicting the availability of TLCF, I determine the methods' percentage of correct TLCF status predictions made. I find that none of the methods examined performs particularly well in predicting the availability of TLCF. For the most accurate method, an ILCF based on firms' EBT from four past years, the percentage of correct predictions is equal to 79.08% only. The method's accuracy is likely to decrease even further in settings where a large number of firms operating in the industries Capital Goods, Consumer Durables & Apparel or Media is employed. In order to make judgements regarding the accuracy of database-driven methods in predicting the amount of TLCF available, I determine the methods' proportion of logarithmic prediction errors that are small in magnitude (where firms' true TLCF amount is either not more than 1.11 times or not more than 0.9 times as large as the predicted amount). I find that the methods examined perform quite poorly in identifying firms' true TLCF amount. For the most accurate method, an ILCF based on firms' EBIT from three past years, the percentage of small prediction errors is equal to 19.64% only. Furthermore, the method's accuracy varies substantially by year of observation.

Overall, my findings indicate that empirical studies relying on database-driven methods might not be able to properly identify firms' TLCF status. Any results obtained regarding

the impact of TLCF are thus likely to be distorted. Previous studies that potentially suffer from accuracy issues due to relying on database-driven identification methods similar to the ones tested here include the investigations of Bernasconi et al. (2005), Overesch and Voeller (2010), Overesch and Wamser (2010), Buettner et al. (2011), Buettner et al. (2012), Dreßler and Overesch (2013), Stöckl and Winner (2013) and Overesch and Wamser (2014). In order to verify the conclusions drawn in these studies regarding the impact of TLCF, I suggest conducting replication studies. These studies should rely on firms' true TLCF status instead of database-driven proxies. My findings further suggest that future studies should not rely on database-driven identification methods at all. In empirical tax research, there is thus a strong need for firm-specific TLCF status information provided by local authorities.

My analysis is a first attempt to close the gap in empirical tax research regarding the accuracy of database-driven identification methods at single-firm level. It can be enhanced and/or extended in various ways. Due to the fact that I employ a sample of listed Italian parent companies, the results derived are likely to be influenced by the Italian institutional setting during the observation period. It would be interesting to know to what extent my results hold if firms located in countries with an institutional setting different from that of Italy were examined. Moreover, the accuracy of my results critically depends upon my ability to reliably identify firms' true TLCF status. The results of the survey I performed suggest that I am able to accurately determine firms' true TLCF status based on IFRS statement information. Nevertheless, the reliability of my findings would certainly be enhanced if I was able to base my study on firm-specific TLCF status information provided by local authorities. Furthermore, the sample I use tends to be small and specific. It would be interesting to know to what extent my results regarding the accuracy of database-driven identification methods hold if a large and more general sample of firms was employed.

References

- Auerbach, A. J., Poterba, J. M. (1987). Tax-Loss Carryforwards and Corporate Tax Incentives. In M. Feldstein (Ed.), *The Effects of Taxation on Capital Accumulation* (305-342). Chicago, IL: The University of Chicago Press.
- Barclay, M. J., Smith, C. W. (1995). The Priority Structure of Corporate Liabilities. *Journal of Finance*, 50, 899-917.
- Bernasconi, M., Marenzi, A., Pagani, L. (2005). Corporate Financing Decisions and Non-Debt Tax Shields: Evidence from Italian Experiences in the 1990s. *International Tax and Public Finance*, 12, 741-773.
- Buettner, T., Overesch, M., Schreiber, U., Wamser, G. (2012). The Impact of Thin-Capitalization Rules on the Capital Structure of Multinational Firms. *Journal of Public Economics*, 96, 930-938.
- Buettner, T., Overesch, M., Wamser, G. (2011). Tax Status and Tax Response Heterogeneity of Multinationals' Debt Finance. *Public Finance Analysis*, 67, 103-122.
- Burke, W. J. (2009). Fitting and Interpreting Cragg's Tobit Alternative Using Stata. *The Stata Journal*, 9, 584-582.
- Cooper, M. G., Knittel, M. J. (2010). The Implications of Tax Asymmetry for U.S. Corporations. *National Tax Journal*, 63, 33-62.
- Dreßler, D., Overesch, M. (2013). Investment Impact of Tax Loss Treatment: Empirical Insights from a Panel of Multinationals. *International Tax and Public Finance*, 20, 513-543.
- Giacometti, P. (2009). Italy Implements Provisions for the Tax Treatment of IFRS Adopters. *International Tax Review*, 20, 59-59.
- Kager, R., Niemann, R. (2013). Income Determination for Corporate Tax Purposes Using IFRS as a Starting Point: Evidence for Listed Companies within Austria, Germany, and The Netherlands. *Journal of Business Economics*, 83, 437-470.
- Kager, R., Niemann, R., Schanz, D. (2011). Estimation of Tax Values Based on IFRS Information: An Analysis of German DAX30 and Austrian ATX Listed Companies. *Accounting in Europe*, 8, 89-123.
- Kinney, M. R., Swanson, E. P. (1993). The Accuracy and Adequacy of Tax Data in COMPUSTAT. *Journal of American Taxation Association*, 15, 121-135.

- Mackie-Mason, J. K. (1990). Do Taxes Affect Corporate Financing Decisions? *Journal of Finance*, 45, 1471-1493.
- Mills, L. F., Newberry, K. J., Novack, G. J. (2003). How Well Do Compustat NOL Data Identify Firms with U.S. Tax Return Loss Carryovers. *Journal of the American Taxation Association*, 25, 1-17.
- Niemann, R., Rechbauer, M. (2013). Wie können Unternehmen mit steuerlichen Verlustvorträgen identifiziert werden? Ergebnisse einer Replikationsstudie [How to Identify Firms with Tax Loss Carry-Forwards? Results of a Replication Study]. *Bankarchiv*, 61, 176-186.
- Overesch, M., Voeller, D. (2010). The Impact of Personal and Corporate Taxation on Capital Structure Choice. *Journal of Public Economics*, 66, 263-294.
- Overesch, M., Wamser, G. (2010). Corporate Tax Planning and Thin-Capitalization Rules: Evidence from a Quasi-Experiment. *Applied Economics*, 42, 563-573.
- Overesch, M., Wamser, G. (2014). Bilateral Internal Debt Financing and Tax Planning of Multinational Firms. *Review of Quantitative Finance and Accounting*, 42, 191-209.
- Petermann, S., Schanz, S. (2013). Latente Steuern auf Verlustvorträge: Empirische Evidenz latenter Steuern auf Verlustvorträge in IFRS-Abschlüssen der DAX-30-Unternehmen in den Jahren 2005 bis 2010 [Deferred Taxes on Tax Loss Carry-Forwards: Empirical Evidence on Deferred Taxes on Tax Loss Carry-Forwards in IFRS Statements of DAX30 Corporations between 2005 and 2010]. *Praxis Internationale Rechnungslegung*, 3/2013, 78-83.
- Stöckl, M., Winner, H. (2013): Körperschaftbesteuerung und Unternehmensverschuldung: Evidenz aus einem europäischen Firmenpanel [Corporate Taxation and Debt: Evidence from a European Firm Panel]. *Journal of Economics and Statistics*, 233, 188-205.

A Supplemental Appendix

A.1 Amadeus Data Items

Table A.1 provides a list of Amadeus data items used for the analyses in this study.

Table A.1: Amadeus Data Items Used

This table shows a list of Amadeus data items used for the analyses in this study.

Variable	Amadeus data item
EBT	Profit/Loss Before Taxation
EBIT	EBIT
EBITDA	EBITDA
firm size	Total Assets
incorporation date	Date of Incorporation
industry affiliation	GICS
pre-tax cashflow	Cashflow + Taxation - Extraordinary Profit/Loss

A.2 Mathematical Derivations

$ILCF_AMT_{i,t-1}^n$ can be derived as follows:

For $j = n$:

$$ILCFAMT_{i,t-j}^n = \begin{cases} |EARN_{i,t-j}| & \text{if } EARN_{i,t-j} < 0 \\ 0 & \text{otherwise} \end{cases} \quad (\text{A.1})$$

For $j = n - 1, \dots, 1$ (given that $n > 1$):

$$ILCFAMT_{i,t-j}^n = \max\{ILCFAMT_{i,t-j-1}^n - EARN_{i,t-j}; 0\} \quad (\text{A.2})$$

where $ILCF_AMT_{i,t-j}^n$ ($ILCF_AMT_{i,t-j-1}^n$) is the stock of ILCF of firm i at the end of year $t - j$ ($t - j - 1$). $EARN_{i,t-j}$ is the amount of IFRS earnings realized in year $t - j$.

$INDPR_{i,t}$ can be derived as follows (Burke, 2009):

$$\begin{aligned}
INDPR_{i,t} &= Pr(TAVB_{i,t} = 1 | \sum_{j=1}^{k-1} INDUSTRY_{i,j}) \\
&= \Phi(\hat{\alpha} + \sum_{j=1}^{k-1} \hat{\beta}_j INDUSTRY_{i,j})
\end{aligned} \tag{A.3}$$

where Φ is the standard normal cumulative distribution function. $\hat{\alpha}$ and $\sum_{j=1}^{k-1} \hat{\beta}_j$ represent the coefficient estimates on the intercept and the industry dummies, as determined in the probit regression shown in Equation 2, respectively.

$INDAMTD_{i,t}$ can be derived as follows (Burke, 2009):

$$\begin{aligned}
INDAMTD_{i,t} &= E[TAMTD_{i,t} | TAVB_{i,t} = 1, \sum_{j=1}^{k-1} INDUSTRY_{i,j}] \\
&= \hat{\alpha} + \sum_{j=1}^{k-1} \hat{\beta}_j INDUSTRY_{i,j} + \hat{\sigma} \cdot \frac{\phi\left(\frac{\hat{\alpha} + \sum_{j=1}^{k-1} \hat{\beta}_j INDUSTRY_{i,j}}{\hat{\sigma}}\right)}{\Phi\left(\frac{\hat{\alpha} + \sum_{j=1}^{k-1} \hat{\beta}_j INDUSTRY_{i,j}}{\hat{\sigma}}\right)}
\end{aligned} \tag{A.4}$$

where $\hat{\alpha}$ and $\sum_{j=1}^4 \hat{\beta}_j$ represent the coefficient estimates on the intercept and the industry dummies, as determined in the truncated normal regression shown in Equation 3, respectively. $\hat{\sigma}$ is equal to the standard deviation of the error term of the truncated normal model. ϕ represents the standard normal density function.

A.3 Industry Affiliation - TLCF Status Predictions

The results of the double-hurdle regression analysis specified in Equations 2 and 3 of Section 3.2.3 are shown in Table A.2. The coefficient estimates of the probit model shown in Column 1 of Table A.2, enable me to derive specifications of $INDPR_{i,t}$ for each of the industry types employed. These are shown in Table A.3. The coefficient estimates of the truncated normal model shown in Column 2 of Table A.2, allow me to derive specifications of $INDAMTD_{i,t}$ for each of the industry types employed. These are shown in Table A.4.

A.4 Industry Affiliation and Firm Age - TLCF Status Predictions

The results of the double-hurdle regression analysis specified in Equations 2 and 3 of Section 3.2.3, including firm age, are shown in Table A.5. The coefficient estimates of the probit model shown in Column 1 of Table A.5, enable me to derive specifications of $INDPR_{i,t}$, adjusted for firm age, for each of the industry types employed. These are shown in Table A.6. The coefficient estimates of the truncated normal model shown in Column 2 of Table A.5, allow me to derive specifications of $INDAMTD_{i,t}$, adjusted for firm age, for each of the industry types employed. These are shown in Table A.7.

Table A.2: Industry Affiliation - Double Hurdle Regression Results

This table shows the double hurdle regression results for the method based on industry affiliation, as represented by firms' four-digit GICS code. The results shown in Column 1 are based on the following probit model: $TAVB_{i,t} = \alpha + \sum_{j=1}^{k-1} \beta_j INDUSTRY_{i,j} + \epsilon_{i,t}$. $TAVB_{i,t}$ is equal to one if firm i has TDCF for deduction in year t and zero otherwise. $INDUSTRY_{i,j}$ is equal to one if firm i operates in industry j and zero otherwise. Industry type Other summarizes industries Energy (1010) and Telecommunication Services (5010). Industry Information Technology (4500) serves as the reference category. $\epsilon_{i,t}$ represents the error term. The results shown in Column 2 are based on the following truncated normal model: $TAMTD_{i,t} = \alpha + \sum_{j=1}^{k-1} \beta_j INDUSTRY_{i,j} + \epsilon_{i,t}$. $TAMTD_{i,t}$ represents firm i 's amount of TDCF available for deduction in year t . For an observation to be included in the truncated normal regression, $TAVB_{i,t}$ has to be equal to one. For both analyses, standard errors are shown in parentheses. They are clustered at firm-level. *, ** and *** correspond to significance levels of 10%, 5% and 1%, respectively.

	(1) Probit Model	(2) Truncated Normal Model
<i>Other (0)</i>	-0.103 (0.596)	59,262 (168,915)
<i>Materials (1510)</i>	-0.103 (0.473)	239,886 (306,488)
<i>Industrials Other (2000)</i>	-0.035 (0.508)	-777,852 (1,118,744)
<i>Capital Goods (2010)</i>	-0.647 * (0.386)	-265,676 (410,292)
<i>Consumer Discretionary Other (2500)</i>	-0.410 (0.469)	109,146 (322,237)
<i>Consumer Durables & Apparel (2520)</i>	-0.345 (0.406)	100,765 (173,452)
<i>Media (2540)</i>	-0.398 (0.437)	220,250 (279,283)
<i>Consumer Staples (3000)</i>	-1.075 ** (0.457)	-1,374,116 (2,106,350)
<i>Health Care (3500)</i>	-0.254 (0.501)	213,435 (281,111)
<i>Utilities (5510)</i>	-0.912 * (0.546)	-366,535 (506,381)
<i>Constant</i>	0.199 (0.295)	-334,567 (536,429)
σ		83,077 (58,935)
N	325	139
Wald statistic	9.71	0.93
Pseudo R ²	0.042	

Table A.3: Industry Affiliation - Specifications of $INDPR_{i,t}$

This table shows specifications of $INDPR_{i,t}$ for each of the industry types employed. $INDPR_{i,t}$ represents firm i 's probability of having TLCF in year t , given its industry affiliation. $INDPR_{i,t}$ is determined by relying on Formula A.3 of Supplemental Appendix A.2, which uses the results of the probit regression analysis shown in Column 1 of Table A.2. Φ corresponds to the standard normal cumulative distribution function. The definition of industry type Other is shown in Table A.2.

	$INDPR_{i,t}$
<i>Other (0)</i>	$\Phi(+0.097) = 0.538$
<i>Materials (1510)</i>	$\Phi(+0.097) = 0.538$
<i>Industrials Other (2000)</i>	$\Phi(+0.164) = 0.565$
<i>Capital Goods (2010)</i>	$\Phi(-0.447) = 0.327$
<i>Consumer Discretionary Other (2500)</i>	$\Phi(-0.210) = 0.417$
<i>Consumer Durables & Apparel (2520)</i>	$\Phi(-0.146) = 0.442$
<i>Media (2540)</i>	$\Phi(-0.199) = 0.421$
<i>Consumer Staples (3000)</i>	$\Phi(-0.876) = 0.190$
<i>Health Care (3500)</i>	$\Phi(-0.055) = 0.478$
<i>Information Technology (4500)</i>	$\Phi(+0.199) = 0.579$
<i>Utilities (5510)</i>	$\Phi(-0.712) = 0.238$

Table A.4: Industry Affiliation - Specifications of $INDAMTD_{i,t}$

This table shows specifications of $INDAMTD_{i,t}$ for each of the industry types employed. $INDAMTD_{i,t}$ represents the amount of TLCF available to TLCF firms, given their industry affiliation. $INDAMTD_{i,t}$ is determined by relying on Formula A.4 of Supplemental Appendix A.2, which uses the results of the truncated normal regression analysis shown in Column 2 of Table A.2. ϕ and Φ correspond to the standard normal density function and the standard normal cumulative distribution function, respectively. The definition of industry type Other is shown in Table A.2.

	$INDAMTD_{i,t}$
<i>Other (0)</i>	$-275,535 + 83,107 \cdot \frac{\phi(\frac{-275,535}{83,107})}{\Phi(\frac{-275,535}{83,107})} = 21,800$
<i>Materials (1510)</i>	$-94,798 + 83,107 \cdot \frac{\phi(\frac{-94,798}{83,107})}{\Phi(\frac{-94,798}{83,107})} = 41,408$
<i>Industrials Other (2000)</i>	$-1,113,241 + 83,107 \cdot \frac{\phi(\frac{-1,113,241}{83,107})}{\Phi(\frac{-1,113,241}{83,107})} = 6,137$
<i>Capital Goods (2010)</i>	$-600,698 + 83,107 \cdot \frac{\phi(\frac{-600,698}{83,107})}{\Phi(\frac{-600,698}{83,107})} = 11,094$
<i>Consumer Discretionary Other (2500)</i>	$-225,619 + 83,107 \cdot \frac{\phi(\frac{-225,619}{83,107})}{\Phi(\frac{-225,619}{83,107})} = 25,313$
<i>Consumer Durables & Apparel (2520)</i>	$-234,006 + 83,107 \cdot \frac{\phi(\frac{-234,006}{83,107})}{\Phi(\frac{-234,006}{83,107})} = 24,654$
<i>Media (2540)</i>	$-114,445 + 83,107 \cdot \frac{\phi(\frac{-114,445}{83,107})}{\Phi(\frac{-114,445}{83,107})} = 38,039$
<i>Consumer Staples (3000)</i>	$-1,709,942 + 83,107 \cdot \frac{\phi(\frac{-1,709,942}{83,107})}{\Phi(\frac{-1,709,942}{83,107})} = 4,020$
<i>Health Care (3500)</i>	$-121,265 + 83,107 \cdot \frac{\phi(\frac{-121,265}{83,107})}{\Phi(\frac{-121,265}{83,107})} = 36,971$
<i>Information Technology (4500)</i>	$-334,837 + 83,107 \cdot \frac{\phi(\frac{-334,837}{83,107})}{\Phi(\frac{-334,837}{83,107})} = 18,638$
<i>Utilities (5510)</i>	$-701,628 + 83,107 \cdot \frac{\phi(\frac{-701,628}{83,107})}{\Phi(\frac{-701,628}{83,107})} = 9,585$

Table A.5: Industry Affiliation and Firm Age - Double Hurdle Regression Results

This table shows the double hurdle regression results for the method based on industry affiliation, as represented by firms' four-digit GICS code, and firm age. The results shown in Column 1 are based on the following probit model: $TAVB_{i,t} = \alpha + \beta_1 AGE_{i,t} + \sum_{j=1}^{k-1} \beta_j INDUSTRY_{i,j} + \epsilon_{i,t}$. The definitions of $TAVB_{i,t}$, $INDUSTRY_{i,j}$ and $\epsilon_{i,t}$ are shown in Table A.2. $AGE_{i,t}$ corresponds to the difference between year t and firm i 's year of incorporation. The results shown in Column 2 are based on the following truncated normal model: $TAMTD_{i,t} = \alpha + \beta_1 AGE_{i,t} + \sum_{j=1}^{k-1} \beta_j INDUSTRY_{i,j} + \epsilon_{i,t}$. The definition of $TAMTD_{i,t}$ is shown in Table A.2. For an observation to be included in the truncated normal regression, $TAVB_{i,t}$ has to be equal to one. For both analyses, standard errors are shown in parentheses. They are clustered at firm-level. *, ** and *** correspond to significance levels of 10%, 5% and 1%, respectively.

	(1) Probit Model	(2) Truncated Normal Model
$AGE_{i,t}$	0.002 (0.004)	452 (1,952)
<i>Other (0)</i>	-0.143 (0.596)	39,660 (182,130)
<i>Materials (1510)</i>	-0.210 (0.527)	211,757 (232,464)
<i>Industrials Other (2000)</i>	-0.050 (0.511)	-772,650 (1,073,730)
<i>Capital Goods (2010)</i>	-0.680 * (0.395)	-267,180 (404,075)
<i>Consumer Discretionary Other (2500)</i>	-0.436 (0.473)	97,251 (281,562)
<i>Consumer Durables & Apparel (2520)</i>	-0.377 (0.413)	92,296 (158,598)
<i>Media (2540)</i>	-0.415 (0.438)	212,924 (250,526)
<i>Consumer Staples (3000)</i>	-1.103 ** (0.460)	-1,369,560 (2,043,199)
<i>Health Care (3500)</i>	-0.261 (0.505)	214,291 (276,285)
<i>Utilities (5510)</i>	-0.917 (0.546)	-367,963 (497,350)
<i>Constant</i>	0.167 (0.306)	-338,185 (532,861)
σ		82,551 (56,092)
N	325	139
Wald statistic	9.91	1.18
Pseudo R ²	0.043	

Table A.6: Industry Affiliation and Firm Age - Specifications of $INDPR_{i,t}$

This table shows specifications of $INDPR_{i,t}$, adjusted for firm age, for each of the industry types employed. The definitions of $INDPR_{i,t}$ and Φ are shown in Table A.3. The definition of industry type Other is shown in Table A.2.

	$INDPR_{i,t}$ (Adjusted for Firm Age)
<i>Other (0)</i>	$\Phi(+0.023 + 0.002 \cdot AGE_{i,t})$
<i>Materials (1510)</i>	$\Phi(-0.044 + 0.002 \cdot AGE_{i,t})$
<i>Industrials Other (2000)</i>	$\Phi(+0.117 + 0.002 \cdot AGE_{i,t})$
<i>Capital Goods (2010)</i>	$\Phi(-0.513 + 0.002 \cdot AGE_{i,t})$
<i>Consumer Discretionary Other (2500)</i>	$\Phi(-0.269 + 0.002 \cdot AGE_{i,t})$
<i>Consumer Durables & Apparel (2520)</i>	$\Phi(-0.211 + 0.002 \cdot AGE_{i,t})$
<i>Media (2540)</i>	$\Phi(-0.249 + 0.002 \cdot AGE_{i,t})$
<i>Consumer Staples (3000)</i>	$\Phi(-0.936 + 0.002 \cdot AGE_{i,t})$
<i>Health Care (3500)</i>	$\Phi(-0.095 + 0.002 \cdot AGE_{i,t})$
<i>Information Technology (4500)</i>	$\Phi(+0.167 + 0.002 \cdot AGE_{i,t})$
<i>Utilities (5510)</i>	$\Phi(-0.751 + 0.002 \cdot AGE_{i,t})$

Table A.7: Industry Affiliation and Firm Age - Specifications of $INDAMTD_{i,t}$

This table shows specifications of $INDAMTD_{i,t}$, adjusted for firm age, for each of the industry types employed. The definitions of $INDAMTD_{i,t}$, ϕ and Φ are shown in Table A.4. The definition of industry type Other is shown in Table A.2.

	$INDAMTD_{i,t}$ (Adjusted for Firm Age)
<i>Other (0)</i>	$-298,517 - 452 \cdot AGE_{i,t} + 82,551 \cdot \frac{\phi\left(\frac{-298,517 - 452 \cdot AGE_{i,t}}{82,551}\right)}{\Phi\left(\frac{-298,517 - 452 \cdot AGE_{i,t}}{82,551}\right)}$
<i>Materials (1510)</i>	$-126,423 - 452 \cdot AGE_{i,t} + 82,551 \cdot \frac{\phi\left(\frac{-126,423 - 452 \cdot AGE_{i,t}}{82,551}\right)}{\Phi\left(\frac{-126,423 - 452 \cdot AGE_{i,t}}{82,551}\right)}$
<i>Industrials Other (2000)</i>	$-1,110,805 - 452 \cdot AGE_{i,t} + 82,551 \cdot \frac{\phi\left(\frac{-1,110,805 - 452 \cdot AGE_{i,t}}{82,551}\right)}{\Phi\left(\frac{-1,110,805 - 452 \cdot AGE_{i,t}}{82,551}\right)}$
<i>Capital Goods (2010)</i>	$-605,349 - 452 \cdot AGE_{i,t} + 82,551 \cdot \frac{\phi\left(\frac{-605,349 - 452 \cdot AGE_{i,t}}{82,551}\right)}{\Phi\left(\frac{-605,349 - 452 \cdot AGE_{i,t}}{82,551}\right)}$
<i>Consumer Discretionary Other (2500)</i>	$-240,927 - 452 \cdot AGE_{i,t} + 82,551 \cdot \frac{\phi\left(\frac{-240,927 - 452 \cdot AGE_{i,t}}{82,551}\right)}{\Phi\left(\frac{-240,927 - 452 \cdot AGE_{i,t}}{82,551}\right)}$
<i>Consumer Durables & Apparel (2520)</i>	$-245,882 - 452 \cdot AGE_{i,t} + 82,551 \cdot \frac{\phi\left(\frac{-245,882 - 452 \cdot AGE_{i,t}}{82,551}\right)}{\Phi\left(\frac{-245,882 - 452 \cdot AGE_{i,t}}{82,551}\right)}$
<i>Media (2540)</i>	$-125,257 - 452 \cdot AGE_{i,t} + 82,551 \cdot \frac{\phi\left(\frac{-125,257 - 452 \cdot AGE_{i,t}}{82,551}\right)}{\Phi\left(\frac{-125,257 - 452 \cdot AGE_{i,t}}{82,551}\right)}$
<i>Consumer Staples (3000)</i>	$-1,707,699 - 452 \cdot AGE_{i,t} + 82,551 \cdot \frac{\phi\left(\frac{-1,707,699 - 452 \cdot AGE_{i,t}}{82,551}\right)}{\Phi\left(\frac{-1,707,699 - 452 \cdot AGE_{i,t}}{82,551}\right)}$
<i>Health Care (3500)</i>	$-123,890 - 452 \cdot AGE_{i,t} + 82,551 \cdot \frac{\phi\left(\frac{-123,890 - 452 \cdot AGE_{i,t}}{82,551}\right)}{\Phi\left(\frac{-123,890 - 452 \cdot AGE_{i,t}}{82,551}\right)}$
<i>Information Technology (4500)</i>	$-338,175 - 452 \cdot AGE_{i,t} + 82,551 \cdot \frac{\phi\left(\frac{-338,175 - 452 \cdot AGE_{i,t}}{82,551}\right)}{\Phi\left(\frac{-338,175 - 452 \cdot AGE_{i,t}}{82,551}\right)}$
<i>Utilities (5510)</i>	$-706,129 - 452 \cdot AGE_{i,t} + 82,551 \cdot \frac{\phi\left(\frac{-706,129 - 452 \cdot AGE_{i,t}}{82,551}\right)}{\Phi\left(\frac{-706,129 - 452 \cdot AGE_{i,t}}{82,551}\right)}$

A.5 Descriptive Statistics

Table A.8 shows descriptive statistics regarding firms' true TLCF status.

Table A.8: Descriptive Statistics

This table shows descriptive statistics regarding firms' true TLCF status. The definitions of $TAVB_{i,t}$ and $TAMTD_{i,t}$ are shown in Table A.2.

(A) $TAVB_{i,t}$	N	%
0	186	57.23%
1	139	42.77%
	325	100.00%

(B) $TAMTD_{i,t} TAVB_{i,t} = 1$	TEUR
mean	23,184
standard deviation	26,297
minimum	34
25%-percentile	5,382
median	15,800
75%-percentile	31,367
maximum	162,049

A.6 Availability of TLCF - Error Analysis

Table A.9 shows how certain firm characteristics and the year of observation influence the probability of making a wrong TLCF status predictions for each method m . The results shown below are based on the probit model specified in Equation 5.

A.7 Amount of TLCF - Error Analysis

Table A.10 shows how certain firm characteristics and the year of observation influence the probability of making a large logarithmic prediction error for each method m . The results shown below are based on a variant of the probit model specified in Equation 5.

Table A.9: Availability of TILCF - Error Analysis

This table shows how certain firm characteristics and the observation year influence the probability of making a wrong TILCF status prediction. The results shown below are based on the following probit model: $PE_{i,t}^m = \alpha + \beta_1 \ln(SIZE_{i,t-1}) + \beta_2 AGE_{i,t} + \sum_{k=1}^{k=1} \gamma_j INDUSTRY_{i,j} + \delta_1 Y11_t + \delta_2 Y12_t + \epsilon_{i,t}$. The definitions of $AGE_{i,t}$, $INDUSTRY_{i,j}$ and $\epsilon_{i,t}$ are shown in Tables A.2 and A.5. $PE_{i,t}^m$ is equal to one if method m wrongly predicts firm i 's TILCF status in year t and zero otherwise. $SIZE_{i,t-1}$ is defined as the amount of total assets at the end of year $t-1$. $Y11_t$ ($Y12_t$) is equal to one in year 2011 (2012) and zero otherwise. Year 2010 serves as the reference category. Standard errors (not shown) are clustered at firm-level. *, ** and *** correspond to significance levels of 10%, 5% and 1%, respectively.

	(1)		(2)		(3)		(4)		(5)		(6)			
	$ILCF_{EBT,n=1}$	APE	$ILCF_{EBT,n=1}$	APE	$ILCF_{EBT,n=2}$	APE	$ILCF_{EBT,n=3}$	APE	$ILCF_{EBT,n=4}$	APE	$ILCF_{EBT,n=1}$	APE	$ILCF_{EBIT,n=2}$	APE
$\ln(SIZE_{i,t-1})$	0.06	0.02	0.07	0.02	0.07	0.02	0.07	0.02	0.08	0.02	0.14 **	0.04	0.16 **	0.05
$AGE_{i,t}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01 *	0.00
Other (0)	-0.20	-0.06	0.00	0.00	0.15	0.04	0.15	0.04	0.13	0.03	-0.53	-0.17	-0.60	-0.18
Materials (1510)	0.75 *	0.23	0.78 *	0.22	0.80	0.22	0.80	0.22	0.79	0.21	0.19	0.06	0.12	0.04
Ind. Other (2000)	0.35	0.11	0.46	0.13	0.46	0.13	0.46	0.13	0.46	0.12	0.27	0.08	0.09	0.03
Cap. Goods (2010)	0.38	0.12	0.57	0.16	0.72 *	0.20	0.72 *	0.20	0.72 *	0.19	0.54	0.17	0.52	0.16
Cons. Discret. Other (2500)	-0.21	-0.06	0.03	0.01	0.19	0.05	0.19	0.05	0.19	0.05	0.15	0.05	0.13	0.04
Cons. Durabl. & App. (2520)	0.69 **	0.21	0.81 **	0.23	0.90 **	0.25	0.90 **	0.25	0.76 *	0.20	0.33	0.10	0.24	0.07
Media (2540)	0.54	0.16	0.78 *	0.23	0.87 *	0.24	0.87 *	0.24	0.86 *	0.23	0.64	0.20	0.55	0.17
Cons. Staples (3000)	0.09	0.03	0.33	0.09	0.48	0.13	0.48	0.13	0.47	0.13	-0.02	-0.01	-0.04	-0.01
Health Care (3500)	0.04	0.01	0.28	0.08	0.25	0.07	0.25	0.07	0.01	0.00	-0.43	-0.13	-0.44	-0.13
Utilities (5510)	-0.44	-0.13	-0.23	-0.07	-0.06	-0.02	-0.06	-0.02	-0.09	-0.02	-0.01	0.00	-0.05	-0.02
Y11 _t	-0.07	-0.02	-0.08	-0.02	-0.15	-0.04	-0.15	-0.04	-0.15	-0.04	-0.04	-0.01	-0.13	-0.04
Y12 _t	0.04	0.01	-0.01	0.00	-0.05	-0.01	-0.05	-0.01	-0.14	-0.04	-0.01	0.00	-0.10	-0.03
Constant	-1.73 **		-2.11 ***		-2.25 ***		-2.25 ***		-2.29 ***		-2.81 ***		-2.95 ***	
N	325		325		325		325		325		325		325	
Wald statistic	19.96		16.53		16.84		16.84		16.95		22.89 *		26.23 **	
Pseudo R ²	0.06		0.06		0.06		0.06		0.06		0.08		0.08	

Table A.9: Availability of TLCF - Error Analysis (continued)

	(7) $ILCF^{EBIT,n=3}$		(8) $ILCF^{EBIT,n=4}$		(9) $ILCF^{EBITDA,n=1}$		(10) $ILCF^{EBITDA,n=2}$		(11) $ILCF^{EBITDA,n=3}$		(12) $ILCF^{EBITDA,n=4}$	
	Coefficient	APE	Coefficient	APE	Coefficient	APE	Coefficient	APE	Coefficient	APE	Coefficient	APE
$\ln(SIZE_{i,t-1})$	0.15 **	0.04	0.16 **	0.05	0.04	0.01	0.07	0.02	0.07	0.02	0.07	0.02
$AGE_{i,t}$	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Other (0)</i>	-0.43	-0.13	-0.47	-0.14	-0.58	-0.19	-0.41	-0.13	-0.25	-0.08	-0.25	-0.08
<i>Materials (1510)</i>	0.21	0.06	0.17	0.05	0.62	0.20	0.88 *	0.28	1.08 *	0.33	1.08 *	0.33
<i>Ind. Other (2000)</i>	0.23	0.07	0.03	0.01	1.05 **	0.34	1.18 **	0.37	1.24 **	0.38	1.24 **	0.38
<i>Cap. Goods (2010)</i>	0.72 *	0.22	0.75 *	0.22	0.62	0.20	0.84 **	0.26	1.02 **	0.31	1.02 **	0.31
<i>Cons. Discret. Other (2500)</i>	0.28	0.09	0.26	0.08	0.51	0.16	0.61	0.19	0.79	0.24	0.79	0.24
<i>Cons. Durabl. & App. (2520)</i>	0.46	0.14	0.23	0.07	0.74 *	0.24	0.91 **	0.28	1.14 **	0.35	1.14 **	0.35
<i>Media (2540)</i>	0.70	0.21	0.61	0.18	0.81 *	0.26	1.04 **	0.32	1.21 **	0.37	1.21 **	0.37
<i>Cons. Staples (3000)</i>	0.10	0.03	0.09	0.03	-0.12	-0.04	0.11	0.03	0.28	0.09	0.28	0.09
<i>Health Care (3500)</i>	-0.32	-0.10	-0.32	-0.09	-0.38	-0.12	-0.14	-0.04	0.03	0.01	0.03	0.01
<i>Utilities (5510)</i>	0.10	0.03	0.07	0.02	-0.16	-0.05	0.05	0.02	0.21	0.06	0.21	0.06
$Y11_t$	-0.16	-0.05	-0.10	-0.03	0.10	0.03	0.01	0.00	0.03	0.01	0.03	0.01
$Y12_t$	-0.17	-0.05	-0.16	-0.05	-0.05	-0.02	-0.07	-0.02	-0.14	-0.04	-0.14	-0.04
<i>Constant</i>	-2.85 ***		-3.02 ***		-1.58 *		-2.08 **		-2.25 **		-2.25 **	
N	325		325		325		325		325		325	
Wald statistic	25.90 **		25.73 **		26.02 **		29.23 ***		32.33 ***		32.33 ***	
Pseudo R ²	0.08		0.09		0.09		0.10		0.11		0.11	

Table A.9: Availability of TLCF - Error Analysis (continued)

	(13) $CLCF^{n=1}$		(14) $CLCF^{n=2}$		(15) $CLCF^{n=3}$		(16) $CLCF^{n=4}$		(17) Industry		(18) Industry/Age	
	Coefficient	APE	Coefficient	APE	Coefficient	APE	Coefficient	APE	Coefficient	APE	Coefficient	APE
$\ln(SIZE_{i,t-1})$	-0.04	-0.01	-0.05	-0.02	-0.05	-0.01	-0.05	-0.01	-0.04	-0.01	0.39 ***	0.13
$AGE_{i,t}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Other (0)	-0.32	-0.10	0.10	0.03	0.08	0.02	0.28	0.08	0.23	0.08	-0.89 **	-0.29
Materials (1510)	1.15 **	0.35	1.48 ***	0.44	1.38 ***	0.40	1.57 ***	0.45	0.27	0.10	-0.46	-0.15
Ind. Other (2000)	0.84 **	0.26	1.25 ***	0.37	1.13 **	0.33	1.34 ***	0.38	0.05	0.02	0.07	0.02
Cap. Goods (2010)	0.63 **	0.20	1.11 ***	0.33	1.06 ***	0.31	1.26 ***	0.36	-0.19	-0.07	0.37	0.12
Cons. Discret. Other (2500)	0.07	0.02	0.49	0.14	0.48	0.14	0.69	0.19	0.04	0.01	0.16	0.05
Cons. Durabl. & App. (2520)	0.81 ***	0.25	1.17 ***	0.35	1.11 **	0.32	1.25 ***	0.35	0.10	0.04	0.15	0.05
Media (2540)	0.70 **	0.22	1.12 ***	0.33	1.12 ***	0.33	1.33 ***	0.38	0.04	0.02	0.08	0.03
Cons. Staples (3000)	-0.04	-0.01	0.36	0.11	0.36	0.11	0.55	0.16	-0.64	-0.24	1.14 ***	0.37
Health Care (3500)	0.09	0.03	0.50	0.15	0.30	0.09	0.26	0.07	0.16	0.06	0.16	0.05
Utilities (5510)	-0.57	-0.18	-0.14	-0.04	-0.16	-0.05	0.06	0.02	-0.46	-0.17	0.44	0.14
$Y_{11,t}$	0.08	0.03	0.04	0.01	-0.06	-0.02	-0.06	-0.02	0.05	0.02	0.15	0.05
$Y_{12,t}$	0.18	0.05	0.09	0.03	0.03	0.01	-0.07	-0.02	-0.01	0.00	0.11	0.03
Constant	-0.61	-0.61	-0.81	0.03	-0.87	0.01	-0.07	-0.02	-0.01	0.00	-4.99 ***	0.03
N	325		325		325		325		325		325	
Wald statistic	31.26 ***		32.00 ***		29.10 **		29.02 **		7.13		9.46	
Pseudo R ²	0.10		0.11		0.10		0.11		0.02		0.04	

Table A.10: Amount of TLCF - Error Analysis

This table shows how certain firm characteristics and the observation year influence the probability of making a large logarithmic prediction error for each method m . The results shown below are based on the following probit model: $LLPE_{i,t}^m = \alpha + \beta_1 \ln(SIZE_{i,t-1}) + \beta_2 AGE_{i,t} + \sum_{j=1}^{k-1} \gamma_j INDUSTRY_{i,j} + \delta_1 Y11_t + \delta_2 Y12_t + \epsilon_{i,t}$. The definitions of $AGE_{i,t}$, $INDUSTRY_{i,j}$ and $\epsilon_{i,t}$ are shown in Tables A.2 and A.5. The definitions of $SIZE_{i,t-1}$, $Y11_t$ and $Y12_t$ can be found in Table A.9. $LLPE_{i,t}^m$ is equal to one if $LPE_{i,t}^m$ is either < -0.5 or $> +0.5$ and zero otherwise. $LPE_{i,t}^m$ corresponds to the difference between the logarithm of the TLCF amount predicted by method m and the logarithm of firm i 's true TLCF amount. Standard errors (not shown) are clustered at firm-level. *, ** and *** correspond to significance levels of 10%, 5% and 1%, respectively.

	(1)		(2)		(3)		(4)		(5)		(6)	
	Coefficient	APE	Coefficient	APE	Coefficient	APE	Coefficient	APE	Coefficient	APE	Coefficient	APE
$\ln(SIZE_{i,t-1})$			-0.32	-0.05	-0.11	-0.02	-0.24	-0.04	-0.25	-0.05	0.11	0.02
$AGE_{i,t}$			-0.01	0.00	-0.01	0.00	0.00	0.00	-0.03	0.00	-0.01 *	0.00
Other (0)			(om.)	(om.)	(om.)	(om.)	(om.)	(om.)	(om.)	(om.)	(om.)	(om.)
Materials (1510)			0.74	0.12	1.12	0.20	-0.17	-0.03	(om.)	(om.)	0.16	0.03
Ind. Other (2000)			(om.)	(om.)	-0.44	-0.08	-0.67	-0.11	(om.)	(om.)	-0.02	0.00
Cap. Goods (2010)			-0.11	-0.02	0.33	0.06	-0.22	-0.04	-0.14	-0.03	-1.48 **	-0.28
Cons. Discret. Other (2500)			(om.)	(om.)	(om.)	(om.)	(om.)	(om.)	-0.23	-0.04	-0.48	-0.09
Cons. Durabl. & App. (2520)			-0.38	-0.06	(om.)	-0.08	-0.21	-0.04	(om.)	-0.09	(om.)	-0.23
Media (2540)			-0.74	-0.12	-0.47	-0.08	-0.97	-0.16	-0.51	-0.09	-1.22 *	(om.)
Cons. Staples (3000)			(om.)	(om.)	(om.)	(om.)	(om.)	(om.)	(om.)	(om.)	(om.)	(om.)
Health Care (3500)			-1.05	-0.17	(om.)	(om.)	(om.)	(om.)	(om.)	(om.)	(om.)	(om.)
Utilities (5510)			(om.)	(om.)	(om.)	(om.)	(om.)	(om.)	-0.16	-0.03	(om.)	(om.)
Y11 _t			(om.)	(om.)	1.07	0.19	0.05	0.01	(om.)	(om.)	-0.32	-0.06
Y12 _t			1.32 ***	0.22	0.73 ***	0.13	0.23	0.04	0.83	0.15	0.77	0.14
Constant			5.19	2.50	2.50	4.35	4.35	4.65 *	4.65 *	4.65 *	0.64	
N			53.00	70.00	70.00	85.00	85.00	85.00	37.00	37.00	71.00	
Wald statistic			11.52	12.11	12.11	4.21	4.21	13.25 *	13.25 *	13.25 *	15.85 *	
Pseudo R ²			0.21	0.15	0.15	0.11	0.11	0.25	0.25	0.25	0.22	

Table A.10: Amount of TLCF - Error Analysis (continued)

	(7) $ILCF_{EBIT,n=3}$		(8) $ILCF_{EBIT,n=4}$		(9) $ILCF_{EBITDA,n=1}$		(10) $ILCF_{EBITDA,n=2}$		(11) $ILCF_{EBITDA,n=3}$		(12) $ILCF_{EBITDA,n=4}$	
	Coefficient	APE	Coefficient	APE	Coefficient	APE	Coefficient	APE	Coefficient	APE	Coefficient	APE
$\ln(SIZE_{i,t-1})$	0.02	0.01	-0.12	-0.04	0.58	0.13	-0.09	-0.02	0.13	0.03	0.08	0.02
$AGE_{i,t}$	-0.01	0.00	0.00	0.00	-0.05 **	-0.01	0.03	0.01	-0.01	0.00	-0.01 *	0.00
<i>Other (0)</i>	0.16	0.04	(om.)	(om.)	(om.)	(om.)	(om.)	(om.)	-0.74	-0.16	0.48	0.11
<i>Materials (1510)</i>	(om.)	(om.)	(om.)	(om.)	(om.)	(om.)	(om.)	(om.)	-0.39	-0.08	0.21	0.05
<i>Ind. Other (2000)</i>	0.43	0.11	-0.16	-0.05	(om.)	(om.)	(om.)	(om.)	(om.)	(om.)	(om.)	(om.)
<i>Cap. Goods (2010)</i>	-0.40	-0.10	-0.40	-0.12	(om.)	(om.)	(om.)	(om.)	(om.)	(om.)	(om.)	(om.)
<i>Cons. Discret. Other (2500)</i>	0.64	0.16	0.51	0.15	(om.)	(om.)	(om.)	(om.)	-0.88	-0.19	0.10	0.02
<i>Cons. Durabl. & App. (2520)</i>	0.75	0.19	(om.)	(om.)	0.91	0.21	-1.29	-0.24	-0.54	-0.12	-0.15	-0.03
<i>Media (2540)</i>	-0.12	-0.03	-0.07	-0.02	(om.)	(om.)	(om.)	(om.)	(om.)	(om.)	-0.41	-0.09
<i>Cons. Staples (3000)</i>	(om.)	(om.)	-0.47	-0.14	(om.)	(om.)	-3.19 ***	-0.60	(om.)	-0.89	-0.40	-0.09
<i>Health Care (3500)</i>	0.02	0.00	0.00	0.00	(om.)	(om.)	(om.)	(om.)	(om.)	(om.)	-0.54	-0.12
<i>Utilities (5510)</i>	(om.)	(om.)	(om.)	(om.)	0.29	0.07	(om.)	(om.)	-0.73	-0.16	(om.)	(om.)
$Y_{11,t}$	-0.71 *	-0.18	-0.25	-0.07	(om.)	(om.)	1.25	0.23	0.23	0.05	-0.54	-0.12
$Y_{12,t}$	-0.50	-0.13	0.12	0.03	(om.)	(om.)	(om.)	(om.)	-0.21	-0.04	0.05	0.01
<i>Constant</i>	1.15		2.36		-4.60		1.76		0.26		0.79	
N	93.00		78.00		13.00 *		20.00		55.00		70.00	
Wald statistic	12.80		3.82		8.70 *		13.45 **		13.33		10.18	
Pseudo R ²	0.11		0.05		0.25		0.20		0.13		0.08	

Table A.10: Amount of TLCF - Error Analysis (continued)

	(13) $CLCF^{n=1}$		(14) $CLCF^{n=2}$		(15) $CLCF^{n=3}$		(16) $CLCF^{n=4}$		(17) Industry		(18) Industry/Age	
	Coefficient	APE	Coefficient	APE	Coefficient	APE	Coefficient	APE	Coefficient	APE	Coefficient	APE
$\ln(SIZE_{i,t-1})$			-0.87 *	-0.13	-0.24	-0.05	-0.09	-0.02	0.20	0.06	0.20	0.06
$AGE_{i,t}$			0.08 *	0.01	-0.02 ***	0.00	0.10 *	0.02	0.00	0.00	0.00	0.00
Other (0)			(om.)		0.74	0.15	(om.)		(om.)		(om.)	
Materials (1510)			(om.)		(om.)		(om.)		(om.)		(om.)	
Ind. Other (2000)			(om.)		0.05	0.01	(om.)		-0.46	-0.13	-0.46	-0.13
Cap. Goods (2010)			(om.)		(om.)		(om.)		0.20	0.06	0.20	0.06
Cons. Discret. Other (2500)			(om.)		0.29	0.06	(om.)		(om.)		(om.)	
Cons. Durabl. & App. (2520)			(om.)		(om.)		(om.)		(om.)		(om.)	
Media (2540)			0.36	0.05	0.16	0.03	-1.42	-0.25	(om.)		(om.)	
Cons. Staples (3000)			(om.)		(om.)		(om.)		(om.)		(om.)	
Health Care (3500)			0.01	0.00	-0.03	-0.01	0.57	0.10	(om.)		(om.)	
Utilities (5510)			-1.53	-0.22	(om.)		(om.)		(om.)		(om.)	
$Y11_t$			-0.64	-0.09	0.30	0.06	-1.67 ***	-0.30	(om.)		(om.)	
$Y12_t$			(om.)		1.16 **	0.23	-1.42 ***	-0.25	(om.)		(om.)	
Constant			10.97 *		3.87		1.81		-1.40	-0.10	-1.40	-0.10
N			29.00		56.00		41.00		49.00		49.00	
Wald statistic			11.61 *		27.29 ***		24.47 ***		1.10		1.10	
Pseudo R ²			0.35		0.16		0.34		0.02		0.02	

Impressum:

Arbeitskreis Quantitative Steuerlehre, arqus, e.V.

Vorstand: Prof. Dr. Ralf Maiterth (Vorsitzender),
Prof. Dr. Kay Blaufus, Prof. Dr. Dr. Andreas Löffler
Sitz des Vereins: Berlin

Herausgeber: Kay Blaufus, Jochen Hundsdoerfer,
Martin Jacob, Dirk Kiesewetter, Rolf J. König,
Lutz Kruschwitz, Andreas Löffler, Ralf Maiterth,
Heiko Müller, Jens Müller, Rainer Niemann,
Deborah Schanz, Sebastian Schanz, Caren Sureth-
Sloane, Corinna Treisch

Kontaktadresse:

Prof. Dr. Caren Sureth-Sloane, Universität Paderborn,
Fakultät für Wirtschaftswissenschaften,
Warburger Str. 100, 33098 Paderborn,
www.arqus.info, Email: info@arqus.info

ISSN 1861-8944