KU LEUVEN

CENTER FOR ECONOMIC STUDIES

DISCUSSION PAPER SERIES
DPS15.10

MAY 2015

How sensitive is the analysis of firm and employment dynamics to longitudinal linkage problems?

Karen GEURTS

Econometrics

Faculty of Economics
And Business

How sensitive is the analysis of firm and employment dynamics to longitudinal linkage problems?

Karen Geurts

April 2015

Abstract

Empirical measures of firm and employment dynamics based on administrative datasets are biased due to missing links in the longitudinal observation of firms. This paper quantifies the bias in a set of widely used measures and evaluates the performance of two linkage methods in reducing bias. We find that a linkage method that builds on the continuity of the firm's workforce is more effective for producing reliable estimates of these measures than a traditional record linking method commonly applied by statistical agencies.

Using improved linkages considerably modifies results regarding the contribution of different classes of firms to employment growth. It almost completely shifts the mass of employment at entry and exit towards the smallest firms. It reduces job creation by entrants and job destruction due to exit by half and reveals a greater importance of established firms for employment growth.

Keywords: Firm dynamics; Job creation and job destruction; Firm-level microdata; Linked employer-employee data: Firm linkage

JEL Classification: C81, J23, L11

K. Geurts : Center for Economic studies, KU Leuven.

I am grateful for the advice of Jo Van Biesebroeck. I am also grateful to the Belgian National Social Security Office (NSSO), in particular to Peter Vets for collaboration on the development of the employee-flow method, and to Statistics Belgium, in particular Youri Baeyens and Antonio Fiordaliso, for providing record linkages created for the Eurostat Structural Business Statistics. I also thank Malcolm Brynin and Alex Schmitt for helpful comments.

1 Introduction

Large-scale administrative datasets in which individual plants or firms are observed over a long period of time are increasingly used to study firm and employment dynamics.[1] The data are an excellent source for examining firm-level determinants of entry, exit and growth (Caves 1998; Dunne et al.1988, 1989; Wagner 2007) and for studying gross creation and destruction of jobs underlying net employment growth (Davis et al. 1996a). A well-known but major problem with these data are so called missing longitudinal linkages. When a firm's administrative identification number changes, for example after a change of ownership or legal form, the firm is observed as an exit and a new entrant instead of as a continuing firm. Such missing links between firm ID numbers have disturbing consequences for empirical measures. They lead to spurious measurement of entry and exit, to misclassifications of employment growth across different age and size classes of firms, and to an overall overestimation of firm and job turnover (Haltiwanger et al. 2013). Changes in the firm structure, such as mergers or split-ups, create additional longitudinal linkage problems and lead to more bias in the measures.

To improve administrative data for statistical and research purposes, statistical agencies in Europe and the U.S. mostly apply traditional record linking methods (Eurostat-OECD 2007; Jarmin and Miranda 2002; Clayton and Spletzer 2009). Missing firm linkages are identified with probabilistic matching techniques and supplementary data sources. In the absence of a reference measurement, it is however unclear how well these methods succeed in avoiding bias in the above mentioned measures. This paper argues and empirically illustrates that an alternative linkage method, which has been applied in several countries (Baldwin et al. 1992; Benedetto et al. 2007), is more effective for improving estimates of these measures. The method traces one key input factor of the firm, the stock of individual employees, from one period to the next to track changes in firm ID numbers and in firm structure. The method, which we here call employee-flow method, relies on linked employer-employee data in which both firms and employees are identified with a unique identification number.

This paper separately applies a traditional and an employee-flow linkage method to an administrative dataset of Belgian employer firms from 2003-2012. In addition, longitudinal firm linkages are constructed that are based on both methods combined. The latter provide a benchmark to compare the performance of the two methods. The aim of this comparison is twofold. First, to quantify how much each of the two linkage methods contribute to reducing bias in a set of widely-used dynamics measures. We

[1] In line with other studies for small countries, the analysis in this paper is carried out at the firm level. This contrasts with the approach in large countries such as the U.S., where the establishment is the preferred level of analysis. In smaller countries like most European ones, the firm is the smallest enterprise unit which benefits from a certain degree of autonomy in decision-making. Establishments of the same firm often merely are located at relatively short distance from each other and there is considerable mobility of workers between them. This would falsely be considered as job creation and destruction in an establishment approach. Longitudinal linkage problems are similar in firm and establishment analysis, but the implementation of record linking methods may differ.

consider entry and exit measures, firm-level growth estimates, and gross job creation and gross destruction rates. The second aim is to evaluate which dynamics measures in particular are more sensitive to missing linkages than others and thus require good quality longitudinal linkages to be accurately estimated.

We find that missing linkages are strongly increasing in firm size. The implication for the empirical measures is that job turnover measures prove to be more sensitive to longitudinal linkage problems than firm turnover, and that dynamics measures by size are biased towards larger firms. The employee-flow method contributes most substantially to correcting bias. Applied independently, it produces estimates of the empirical measures that are close to the results based on both methods combined. The traditional method reduces bias only halfway. The bias reduction achieved by applying improved longitudinal linkages is not marginal. Empirical measures based on the employee-flow method markedly differ from the ones obtained by the traditional method, and lead to considerably different conclusions about the contribution of various classes of firms to employment growth. We highlight three findings.

First, missing linkages introduce a strong upward bias in aggregate job reallocation measures. At entry and exit, job creation and job destruction rates are overestimated by more than 80 percent. The traditional method leaves an upward bias of 50 percent. The employee-flow method reduces bias to less than 10 percent and reveals that the employment shares of entrants and exits are actually very low (1.5% and 1.6% respectively).

Second, inadequately edited data falsely suggest that firms entering the market with more than 10 employees account for almost half of job creation at entry, and that more than half of job destruction by exit is due to medium and large firms exiting the market. Applying improved linkages almost completely shifts the mass of employment at entry and exit towards the smallest firms. The employee-flow linkages indeed reveal that most medium and large entrants and exits in the administrative register are brought about by ID changes or firm restructurings. The traditional method only identifies about half of them. Distinguishing these spurious entrants and exits from *de novo* entrants and exits by closure has implications for firm-level analysis that reach beyond the set of empirical measures considered in this paper. Several studies have highlighted that both populations fundamentally differ in characteristics other than size (Baldwin and Gorecki 1987; Acs and Audretsch 1989; Storey 1991; Mata et al. 1995; Geroski 1995; Geurts and Van Biesebroeck 2014).

Finally, misclassifying continuing firms as exits not only disturbs the true size distribution of the exit population, but also leads to an underestimation of the performance of successful firms. In size classes over 5 employees, the employee-flow method leads to estimates of the firm-level growth rate which are 2 percentage points higher than the ones based on the traditional method. If employment growth is expressed in absolute terms, the impact is most substantial for larger firms. The improved measures that are obtained for the Belgian private sector challenge the common perception that small firms are the engine of job creation. Instead they suggest that large established firms contribute a great deal more to employment growth than smaller ones.

For research on job flows and firm dynamics, the employee-flow method has other attractive features. First, while the traditional method deduces information on longitudinal firm histories from a complex set of partial firm characteristics, the employee-flow method directly implements an economic definition of firm continuity. Continuity of one of the firm's key production factors, the stock of employees, is used to identify firms that operate continuously but change identification number or firm structure. Second, the method effectively captures changes in firm structure, such as mergers, take-overs and split-offs. Researchers often do not simply want to neutralize these events, as we do in this paper, but to treat them as events of economic importance.[2] A third advantages lies in international comparability. Traditional link methods depend on supplementary data sources which differ widely across countries. The employee-flow method, by contrast, is based on linkage algorithms which follow general rules and can be standardized across countries.

This paper is organized as follows. Section 2 provides a brief overview of the literature on longitudinal linkage problems and the solutions offered so far. Section 3 describes the data used in this paper and provides the technical background on the two record linking methods that are adopted. It is explained how linked firms are reclassified, and a simple approach is proposed for imputing employment growth of these firms. Section 4 provides background statistics on the linked firms which facilitate interpretation of the results. Section 5 discusses the results. Empirical measures based on the traditional and the employee-flow method are compared with the benchmark results. We consider entry and exit characteristics, firm-level growth estimates, and the mean and annual variance of gross job creation and job destruction rates. Section 6 concludes.

2 Longitudinal linkage problem and solutions

The causes of longitudinal linkage problems in firm-level administrative datasets and the consequences for the study of firm and employment dynamics are well known (Baldwin et al. 1992; Acs and Armington 1998; Haltiwanger et al. 2013). A first problem stems from changes in the administrative firm identification number. A new ID number may be assigned by the administration when the ownership or legal form changes, or firms themselves may transfer their activities into a newly registered company for tax optimization or liability avoidance. Missing links between old and new ID numbers generate

[2] Employee-flow methods have been used to identify different types of firm restructurings or to examine a specific group of entrants (see for example Persson 2004; Eriksson and Kuhn 2006; Benedetto et al. 2007). For other applications of linked employer-employee data see Abowd and Kramarz (1999).

‘spurious’ entrants and exits which introduce an upward bias in measures of firm and employment turnover (Jarmin and Miranda 2002). Changes in the firm structure brought about by mergers, takeovers or split-offs create additional longitudinal linkage problems. They lead to creations and closures of firm identification numbers that are clearly different from *de novo* firm entry and exit by failure (Dunne et al. 1988; Baldwin and Gorecki 1987). They also lead to administrative transfers of employees between ID numbers which appear as shocks to firm-level employment in the data. This further inflates job reallocation measures (Pinkston and Spletzer 2002). Overestimation of aggregate measures of entry, exit, job creation and job destruction is one empirical problem. Several authors have pointed out that missing longitudinal linkages also cause distortions in firm-level measurements. They introduce a size bias in firm-level estimates of entry, exit and growth (Geurts and Van Biesebroeck, 2014), lead to misclassifications of the firm age (Haltiwanger et al. 2013), and hamper comparative analysis of firm demographics (Bartelsman et al. 2005).

In response to these concerns, statistical agencies have developed longitudinal business databases for research, mostly relying on traditional record linking techniques to identify missing links between firm identification numbers.[3] These methods primarily make use of supplementary data sources with information on firm demography, ownership changes or M&A activity. Although such sources provide valuable additional information, firm changes that are not registered remain out of scope. Therefore, the link procedures are usually complemented by probabilistic matching techniques which exploit similarities in partial firm identifiers, such as name, address, or industry code to establish links between ID numbers of the same firm. The U.S. Census Bureau has taken this approach one step further for the creation of a longitudinal establishment and enterprise database (Acs and Armington 1998; Jarmin and Miranda 2002). The longitudinal histories of the firm establishments are used to identify true firm entry and exit and changes in firm structure. Establishments that change identification number are in turn linked by a probabilistic matching procedure. In the same spirit, Mata et al. (1995) and Baldwin and Gorecki (1987) use the longitudinal identifier of the parent firm to distinguish between *de novo* planet entry and plants created by already established firms.

At the same time, several countries have developed an employee-flow method for the improvement of longitudinal firm linkages (Baldwin et al. 1992; Benedetto et al. 2007).[4] The method takes an entirely

[3] Clayton and Spletzer (2009) describe the longitudinal linkage method currently applied by the U.S. Bureau of Labor Statistics. In Europe, linkage methods adopted by national statistical agencies have led to general Eurostat-OECD recommendations on firm record linking (Eurostat-OECD 2007). The recommended methods aim at the construction of harmonised statistical indicators on firm demography, such as the Eurostat Structural Business Statistics and the OECD Entrepreneurship Indicators Programme.

[4] One of the first institutes to implement an employee-flow method has been Statistics Canada (Baldwin et al. 1992), where it is still used for the construction of the National Accounts Longitudinal Microdata File (Rollin 2013). Employee-flow methods are also used for the construction of longitudinal employer databases in Denmark (Albaeck and Sorensen 1998), Finland (Maliranta and Nurmi 2004), Sweden (Persson 2004) and Germany (Hethey and Schmieder 2013).

different approach to identify missing linkages. While traditional methods retrieve information on firm continuity from a complex set of partial firm characteristics, employee-flow methods use one key input factor of the firm, the workforce, to trace the firm's longitudinal history and changing structure over time. If a firm changes identification number while continuing its operations, one of the main production factors, the stock of individual employees, is likely to remain largely the same. Continuity of the workforce from one period to the next could then be used to detect changes in firm ID numbers. Similarly, when firms are merged, split up or parts are sold to another firm, this will be reflected in a merge or division of workforces. Employee-flow methods make use of linked employer-employee data to implement this workforce-based characteristic of firm continuity. Large clusters of employees that appear to 'move' from one firm identification number to another are used to signal changes in ID numbers or in the firm structure.

From a methodological viewpoint, it is clear that both linkage methods have strengths and weaknesses. The employee-flow method uses an economically meaningful definition of firm continuity that surpasses the limitations of the administrative identification number. This definition is translated in linkage algorithms that are independent of legal regulations and can be standardized across countries. The method, however, is unsuited to capture links among the smallest firms, as will be discussed below. Traditional link methods do cover all size classes, but partly depend on country-specific notions of firm continuity and on the type of information that can be retrieved from supplementary data sources. Probabilistic matching helps to identify additional linkages, but major changes in discriminating identifiers, e.g. name or telephone number, strongly reduce the probability of a positive match.

From an empirical viewpoint, the strengths and weaknesses of the two methods are less obvious. Both traditional and employee-flow methods are found to remove a substantial amount of 'spurious' firm and job turnover from the data (Jarmin and Miranda 2002; Pinkston and Spletzer 2002; Benedetto et al. 2007). Empirical dynamics measures based on improved longitudinal data thus more accurately reflect the true dynamics in the economy. In the absence of comparison, the reliability of the revised datasets is generally taken for granted by researchers. However, to our knowledge, no serious attempt has been undertaken so far to evaluate bias in empirical measures after implementation of the record linking methods, or to determine which linkage method is more effective in avoiding bias in the measures.

3 Data and methods

The register of Belgian employer firms that is used in this paper is particularly well suited to investigating these questions. First, the two record linking methods that we apply to this dataset are illustrative examples of the traditional and the employee-flow approach. The traditional method was developed by

Statistics Belgium in line with the OECD-Eurostat recommendations on firm record linking (Eurostat-OECD 2007). These guidelines aim at the construction of harmonized business registers and statistical indicators on firm demography. The employee-flow method was developed by the National Social Security Office in collaboration with the University of Leuven (Geurts and Vets 2013) and builds on similar examples in other countries, in particular Canada (Baldwin et al. 1992), Sweden (Persson 2004), and the U.S. (Benedetto et al. 2007).[5] Further, the initial longitudinal firm linkages present in the Belgian firm register are relatively consistent: firm identification numbers generally do not change after a change in ownership or legal form, while this is one of the main reasons for longitudinal linkage errors in many other countries. The benchmark results obtained after implementation of both linkage methods combined can therefore be considered fairly accurate estimates of firm and employment dynamics in the economy.

Identifying longitudinal firm histories and, by extension, the point of entry and exit, requires an operational definition of entrants and exits. In this paper, 'real' entrants are defined as firms that enter the market by starting new operations and creating new employment positions. Likewise, real exits correspond to firms that shut down and terminate all existing employment contracts. In between, firms are defined as continuing, also when they merge or split up activities. This definition of entrants closely corresponds to the concept used in theoretical models of firm dynamics (Jovanovic 1982), as well as to the definition of entrants that is implicitly assumed in most empirical studies (Caves 1998). Such *de novo* entry is opposed to entry by established firms which can take a variety of forms (Caves and Porter 1977). The definitions of entry and exit used in this paper are also in line with the ones recommended by OECD and Eurostat, as will be discussed below.

3.1 Linked employer-employee dataset

The register of Belgian employer firms is maintained by the Belgian National Social Security Office (NSSO) and is based on quarterly social security declarations. It covers all private firms with at least one employee in the period from 2003-2012, including 200,000 active firms and 2,500,000 employees on average per year.[6] Table A.1 in the Appendix reports the number of firms and employees in the dataset, classified into eight industry groups.

The register is a linked employer-employee dataset. Both employers and employees are identified by means of a unique identification number. This information is exploited for the employee-flow method. The NSSO employer number is uniquely linked to the official firm identification number, the CBE

[5] The methods has been developped for the construction of the DynaM longitudinal employer database. The database is designed to track changes in employment both at the macro and the firm level and to produce annual series of gross job gains and losses statistics in Belgium. See www.dynam-belgium.org

[6] Temporary work agencies are left out from the analysis in this paper because the high job turnover in this industry confuses the discussion of average job reallocation.

number. The CBE number, which is used by all government administrations, enables us to implement results of the traditional record linking method into the NSSO dataset. The CBE/NSSO number ensures good quality longitudinal firm records. Upon registration, new firms receive a CBE number which they keep for their entire lifetime. Unlike in many other countries, the firm identification number is unaltered in the case of a change of ownership or legal form.

A new CBE number is however assigned by the administration in a few situations: when a self-employed individual turns his or her business into a company, and when a firm changes ownership after bankruptcy. Furthermore, firm-induced ID changes occur for similar reasons as it is the case in other countries. For the purpose of accounting advantages or avoidance of liability, firms may exit by voluntary liquidation or bankruptcy and continue the same activities in a newly registered company.

As in other countries, firm identification numbers may also be created or disappear when legal entities are merged or split-up. Such changes in the firm structure may reflect actual mergers, acquisitions, break-ups or divestitures, but also mere administrative transfers of activities between legal units of the same controlling enterprise. A common example of the latter is the subdivision of the firm in smaller entities with separate firm identification numbers. For expanding firms, this is a way to remain below the size thresholds for legal obligations.[7] Other reasons why firms create different ID numbers are tax advantages (the separate entities are not considered as part of the same firm) or limitation of liability. The practice is common under the form of enterprises controlling a network of local affiliates, but it is also used to distribute the firm activities across different industries with differential regulations.

3.2 Record linking methods

3.2.1 Traditional record linking method

The traditional linkage method that is applied in this paper relies on probability-based matching and the use of supplementary data sources. It has been developed by Statistics Belgium within the Eurostat-OECD framework on business demography. Eurostat and OECD provide clear-cut definitions of enterprise ‘births’ and ‘deaths’ (Eurostat-OECD 2007). Firm identification numbers that enter and exit for other reasons should be filtered out. A birth, for instance, “amounts to the creation of a combination of production factors with the restriction that no other enterprises are involved in the event.” Births should not include entrants due to restructurings of a set of enterprises such as mergers or break-ups, newly created enterprises after a change of legal form, take-overs of the activity of an existing enterprise, creations of additional legal units solely for the purpose of providing a single production factor or an

[7] In Belgium, small firms do not need to file full annual accounts or install a works council (with fewer than 100 employees, turnover below 7.3m EUR, and balance sheet total below 3.65m EUR).

ancillary activity, and so on. Likewise, a death "coincides with the dissolution of a combination of production factors with the restriction that no other enterprises are involved in the event".

Statistics Belgium uses information from Commercial Court files and from the NSSO to identify changes in firm ID numbers and in firm structure. The Commercial Court provides information on official mergers, acquisitions and split-ups, and on changes in the CBE number. The link between the NSSO and the CBE number further help to track firms that change ID number.

The linkages are complemented with a probabilistic matching procedure. Similarities in name, address, and 4-digits industry code are used to compute probabilities that records refer to the same firm. Automatic and industry-specific ad-hoc rules are applied to verify the results. Eventually, all accepted matches, as well as an important part of rejected and probable matches, are validated by clerical review. Although advanced software is adopted to minimize false (non-)matches[8], probability-based matching of firm records is subject to subjective evaluation and analyst intervention. Major changes in discriminating identification numbers, e.g. name or telephone number, reduce the probability of a positive match, while such modifications often occur at the very moment a firm implements a legal or organizational change. Probabilistic matching is also less suitable for the identification of changes in firm structure such as mergers and split-ups.

3.2.2 Employee-flow method

Employee-flow methods use one main criterion to establish linkages between firm identification numbers: similarity of the workforce. Actual implementations of this method are basically similar in design. Changes in firm ID numbers and in firm structure are identified by tracing large clusters of employees that appear to 'move' from one firm identification number to another between two subsequent observations in time. The methods rely on the assumption that the simultaneous transition of a significant number of employees from one firm identification number to another is unlikely to be the result of individual worker mobility. Therefore, actual linkage procedures generally start from a minimum cluster of three to five employees, as for smaller clusters, there is a high probability that the employee flow merely represents individual job changes. This absolute threshold is supplemented with a set of relative thresholds, which aim at avoiding false matches and at distinguishing between different types of firm restructurings. Due to the minimum cluster size, employee-flow methods are inappropriate for linking small firms. Yet they do achieve high coverage of linkages between larger firms, where sufficiently large clusters of employees can be followed over time.

The employee-flow linkages applied in this paper are generated by a simple linkage algorithm that consists of two stages. In a first stage, clustered employee flows are selected that include at least five

[8] The matching procedure used by Statistics Belgium is based on the Term Frequency – Inverse Document Frequency method.

individual employees which 'move' from one ID number (the 'predecessor') to another (the 'successor') between two quarterly observations.[9] The simultaneous transition of a significant number of employees between ID numbers in such a short time span is a first indication that the employee flow is not the result of individual job changes. The second stage consists of a set of decision rules that aim to capture different forms of inter-firm linkages. The rules set thresholds for the relative cluster sizes, i.e. the size of the clustered employee flow relative to the total workforce of the firms involved. Section B.1 in the Appendix describes the full set of rules and their formal conditions. Three major rules cover 90 percent of all linkages and are briefly discussed below.

The first decision rule covers the major part of employee-flow linkages (57%) and captures links between ID numbers with largely identical workforces. Two firm identification numbers are linked if the employee-flow cluster represents at least 50 percent of the workforce of both the predecessor and the successor. This condition is a formal translation of our workforce-based definition of firm continuity: two successive firm identification numbers that employ *mostly* the same workforce, are considered to refer to the same firm.

Two other major rules identify links between smaller and larger firms. A firm may disappear from the dataset while continuing its activities as part of a larger entity. Such 'absorptions' do not meet our definition of exit, and the transfer of workers to the larger entity does not correspond to the destruction and creation of jobs. To capture these events, a link is established if at least 75 percent of the workforce of an exiting firm is transferred to an already established firm. This second rule identifies 22 percent of additional linkages. The third rule captures the opposite case, when part of the activities of a continuing enterprise is transferred to a new firm ID number. If the employee cluster coming from the established firm represents at least 75 percent of the workforce of the new entrant, a link between the two ID numbers is established. Such 'split-offs' cover an additional 11 percent of linkages. Mergers, break-ups and more complex forms of inter-firm linkages are identified with other decision rules described in the Appendix B.1. They each cover only small parts of additional linkages.

Several robustness checks are performed to test the sensitivity of the empirical results to the set of conditions imposed in the linkage algorithm. Relaxing or restricting the relative cluster size thresholds has little impact on the results. Reducing the set of decision rules has little impact either. The robustness results are discussed in the Appendix B.2.

3.2.3 Both record linking methods combined

We also construct longitudinal firm linkages that incorporate all information provided by both the traditional and the employee-flow method. Linkages edited in this way are the most accurate longitudinal

[9] This absolute minimum threshold is in line with other recent applications of the method (Benedetto et al. 2007; Rollin 2013).

firm records that can be obtained with the available methods. They are used to calculate empirical benchmark measures to which results obtained by the two individual methods are compared.

3.3 Recalculating measures of entry, exit and growth

Measures of firm and employment dynamics presented in this paper are computed as year-by-year changes between June 30th of year *t-1* and year *t*. The entry and exit of a firm are defined as the first and last year it reports positive employment.[10] In between, firms are labeled as continuing. Continuing firms may have no employees in a given year.

Improved longitudinal linkages are primarily used to identify continuing firms which are misclassified as entrants and exits. They will be labeled as 'spurious' entrants and exits. As is the common practice, they are removed from the entry and exit populations. Recalculating firm-level growth measures is more challenging, as several firms can be interlinked in a given period. To our knowledge, no satisfactory solution has been suggested so far. We propose a simple solution for imputing employment growth at the firm level. Aggregate statistics then follow naturally from the revised firm-level observations.

The following example illustrates the problem. Suppose a link is identified between two firms that merge into a new administrative entity. The two firms, previously misclassified as exits, are now identified as continuing. The jobs of these firms are not lost, neither should the jobs that are transferred to the new entity be treated as job creation. However, job growth or decline at the aggregate level of both firms does reflect true job creation or destruction.

The approach adopted here is to construct an aggregate event level including all firms interlinked in a given period from *t-1* to *t*. Firm-level employment in *t* is imputed by assuming the same growth rate for each firm involved in the event. The advantage of this approach compared to solutions proposed elsewhere is that we do change the firm counts and preserve the firm size distribution at the beginning of each period.[11] This allows for a direct comparison of firm-level growth rates before and after the linkage procedures. The imputation procedure has a straightforward interpretation in the case of most types of events, as is discussed in the Appendix C.

Imputation of employment is performed on a year-by-year basis, i.e. for firms involved in an event in a given period *t-1* to *t*. In the next period, we restart from registered employment in *t* and impute

[10] Since the dataset covers the period 2003-2012, entry and exit can be defined for nine annual periods: from 2003-2004 to 2011-2012.

[11] The approach of the U.S. Bureau of Labor Statistics (Pinkston and Spletzer 2002) is to collapsing both firms into an aggregated employer and calculate employment change at the level of this aggregate entity. For aggregate measures of job creation and destruction, our strategy yields the same result as this approach. The disadvantage of the BLS approach is that the firm counts will be inconsistent across time and, more importantly, the relation between firm size and firm growth will be biased. Indeed, the size of the aggregate employer is by construction larger than those of the original firms.

employment in *t+1* for events in that period. Geurts and Van Biesebroeck (2014) have extended the imputation method over a five-year period. They found that firm ID numbers involved in a linkage event are more likely to be involved in another event in one of the following years. Reconstructing longitudinal employment histories of firms over several periods thus quickly turns into a complex exercise which has to take into account multiple events of interlinked identification numbers.

4 Characteristics of linked firms

To facilitate understanding of the impact of the linkage methods on the empirical measures, this section provides some relevant background statistics on linked firms.

4.1 Events leading to missing linkages

Table 1 summarizes the types of events that give rise to spurious entrants, spurious exits, and linked continuing firms. We use a basic classification of events that clarifies the main reasons behind missing linkages. The results are reported for each linkage method separately.

Table 1 Types of events leading to spurious entrants, spurious exits and linked continuing firms

	Type of event					
a. Spurious entrants						
Linkage method:	ID change	Split-off, break-up	Merger	Combination	Total	
Traditional method	57	36	1	6	100	*(n=1149)*
Employee-flow method	65	30	2	3	100	*(n=952)*
Both methods combined	57	35	1	7	100	*(n=1867)*
b. Spurious exits						
Linkage method:	ID change	Absorption, merger	Break-up	Combination	Total	
Traditional method	45	49	0	6	100	*(n=1469)*
Employee-flow method	54	42	1	3	100	*(n=1163)*
Both methods combined	49	43	1	7	100	*(n=2207)*
c. Linked continuing firms						
Linkage method:	Absorption	Split-off	All continuing	Combination	Total	
Traditional method	12	9	77	3	100	*(n=8630)*
Employee-flow method	46	31	13	11	100	*(n=944)*
Both methods combined	14	11	71	4	100	*(n=9240)*

Note: Annual averages over the 2003-2012 period.

Spurious entrants mostly emerge from ID changes. An ID change is defined as a one-to-one link between a firm that continues operations with a new ID number without other firms being involved in the event. ID changes explain more than half of the spurious entrants identified by the traditional method (57%), and two thirds of the ones identified by the employee-flow method (65%). Split-offs of parts of firms or full break-ups into new entities give rise to about another third of misclassified entrants in both linkage methods. Only few spurious entrants originate from mergers or more complex events. The share of this last category slightly increases if linkages from both methods combined are taken into account.

Spurious exits largely result from ID changes as well, as counterparts of spurious entrants. In addition, more than 40 percent of misclassified exits are explained by firms that are absorbed by an established firm or merged with other exits into a new firm ID number.

The number of continuing firms that are linked to another firm identification number strongly differs between the linkage methods. The employee-flow method mainly identifies established firms that take over the workforce of an exiting ID number (counterpart of spurious exit due to absorption), or split off part of their activities into a new legal entity (counterpart of spurious entrant due to split-off). The traditional method captures many more links, especially between two continuing firms. The latter predominantly refer to links between large conglomerates of legal entities that are part of the same controlling company, such as in retail. The probabilistic matching procedure re-identifies these links in each period. However, as long as no restructuring of activities between the entities occurs, such links have little impact on the empirical dynamics measures, as will be shown below.

4.2 Spurious entrants and exits by size

Table 2 provides the main explanation for the bias in the empirical measures discussed in the next section. It presents the percentage shares of administrative entrants and exits that are identified as spurious ones by either of the linkage methods. The first column reports the shares in the total population of entrants or exits, the other columns give the shares in a given size class. The benchmark results based on both linkage methods combined show that one in ten administrative entrants and one in eight exits are identified as spurious. Both the traditional and the employee-flow method capture a much lower share of misclassified firms, which indicates a high degree of complementarity between the two methods.

Two important patterns emerge from Table 2. First, the probability that a new firm identification number corresponds to a spurious entrant increases dramatically with size, and the same holds for exits. This pattern is consistent across both link methods. The traditional linkage method identifies about 5 percent of the smallest entrants and exits as being misclassified, which amounts to more than 40 percent in the largest size class. The employee-flow linkages reveal this pattern even more sharply. It shows that one in three entrants and exits with 5 to 9 employees and almost all entrants and exits with over 100 employees are brought about by ID changes or firm restructurings.[12] The implication for the entry and exit measures is that missing links will have a larger effect on the employment shares of entrants and exits than on the firm entry and exit rates.

Table 2 Share of spurious entrants and spurious exits by size

		Firm size (number of employees) at entry or exit					
	Total	1-4	5-9	10-19	20-49	50-99	100+
Entrants							
Unedited data *(n)*	*19069*	*16852*	*1345*	*527*	*255*	*54*	*37*
% identified as spurious:							
Traditional method	6	5	13	16	20	31	41
Employee-flow method	5	-	30	52	67	77	97
Both methods combined	10	5	36	55	70	80	97
Exits							
Unedited data *(n)*	*18692*	*16058*	*1454*	*649*	*374*	*96*	*60*
% identified as spurious:							
Traditional method	8	6	16	23	33	42	57
Employee-flow method	6	-	30	50	65	75	90
Both methods combined	12	6	36	54	69	78	91

Note: Annual averages over the 2003-2012 period.

The second observation from Table 2 is that the added value of the traditional method in medium and large size classes is rather low. In all size classes above 5 employees, the employee-flow method captures two to three times more spurious entrants and exits than the traditional method, and closely approximates the results obtained when using both methods combined. The traditional method is however necessary for identifying misclassified firms in the smallest size class (1-4 employees), where employee-flow linkages are absent by construction. The close approximation between the employee-flow and benchmark results in larger size classes will be reflected throughout all empirical measures with an employment component. The employee-flow method indeed reveals that most medium and large entrants and exits in the administrative dataset are brought about by ID changes or firm restructurings.

It will be shown below that correctly identifying entry and exit has a dramatic impact on job reallocation measures, as spurious entrants and exits represent important shares at entry and exit. An accurate distinction between what we have labelled as *real* versus *spurious* entrants and exits has however implications for firm-level analysis that reach far beyond the set of measures considered in this paper.

[12] Table 2 highlights that newly registered firms with over 50 employees are most likely incumbents that continue operations – either in total or partially – with a new identification number. Likewise, if a large firm exits the dataset, there is a high probability that it refers to a continuing employer that has transferred its activities to another legal entity. The observation that firms entering the market with over 50 or 100 employees are exceptional is intuitive but often not reflected in administrative registers. Datasets with quite a few large entrants should alert the researcher, as he or she is most probably mistaking established firms for entrants.

Earlier studies that have made a similar distinction have found pronounced differences between the two types of entrants and exits. Treating them as a homogeneous group can lead to highly misleading conclusions about entry and exit characteristics. One distinctive feature is size. Baldwin and Gorecki (1987) and Mata (1993) have shown that entry by established firms and exits brought about by changes in the firm structure are many times larger than *de novo* entrants and exits by closure. The results in this paper confirm these findings. Spurious entrants and exits are eight times larger on average than real entrants and exits (Table A.2 in the Appendix). Other studies have demonstrated that entry and exit from ID changes, firm restructurings or diversifying firms also differ in characteristics other than size, such as in the determinants of entry (Acs and Audretsch 1989; Storey 1991), post-entry growth patterns (Mata et al. 1995; Geurts and Van Biesebroeck 2014), or the profitably and productivity at exit (Baldwin and Gorecki 1987).

4.3 Firm with imputed employment growth

Firm-level employment growth of linked firms is recalculated with the imputation method described in Section 3.3. The revision applies to spurious exits reclassified as continuing firms and to continuing firms that are linked to another ID number. A moderate share of firms are affected and, although substantial, the impact on the empirical measures will be less than the impact of filtering out spurious entrants and exits.

Table A.3 in the Appendix reports the share of firms of which employment growth is considerably revised after the imputation procedure, i.e. a relative adjustment of more than 10 percent compared to the unedited data. The benchmark results based on both link methods combined show that on average 3.8 percent of active firms in a given period are concerned. This share increases with size and amounts to more than 7 percent of the firms with over 50 employees. The traditional link method affects more firms than the employee-flow method, especially in size classes under 50 employees. The employee-flow method, however, has a greater impact on larger firms, which will be reflected in a more substantial revision of total job reallocation measures.

5 Results

This section discusses the sensitivity of empirical measures of firm and employment dynamics to missing links in longitudinal firm histories. The measures are evaluated before and after implementing either of the linkage methods, and compared with benchmark results based on both methods combined. All measures are computed as year-by-year changes between June 30th of year *t-1* and year *t*.

5.1 Entry and exit dynamics

In the previous section, it has been shown that the shares of spurious entrants and exits increase with firm size. The implication for the entry and exit patterns is that missing links have a larger effect on the employment shares of entrants and exits than on the firm entry and exit rates.

Table 3 Summary statistics of entry and exit

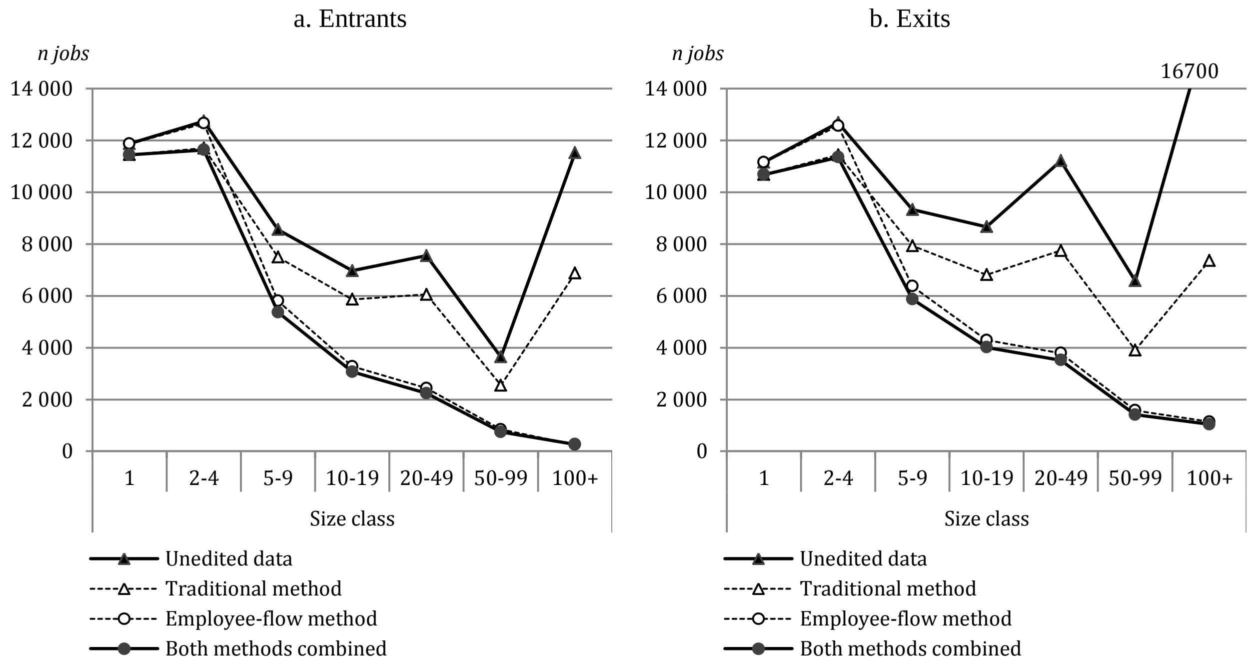
	Entry measures			Exit measures		
	Entry rate (%)	Job creation rate (%)	Average size (employees)	Exit rate (%)	Job destruction rate (%)	Average size (employees)
a. Measures by linkage method						
Unedited data	9.6	2.5	3.3	9.4	3.1	4.1
Traditional method	8.8	2.1	2.9	8.6	2.2	3.1
Employee-flow method	9.1	1.5	2.0	8.8	1.6	2.3
Both methods combined	8.7	1.4	2.0	8.3	1.5	2.3
b. Percent bias vs. both methods combined						
Unedited data	*11%*	*81%*	*64%*	*13%*	*102%*	*78%*
Traditional method	*2%*	*50%*	*46%*	*4%*	*48%*	*35%*
Employee-flow method	*5%*	*7%*	*1%*	*6%*	*8%*	*2%*

Note: Annual averages over the 2003-2012 period. The entry (exit) rate represents the share of entrants (exits) in all active employers of a given period. The job creation (job destruction) rate represents the employment share of entrants (exits).

Table 3 indeed shows that entry and exit rates are only moderately revised downwards after applying the linkage procedures. The results in the first row are based on the unedited administrative data. The next rows show the revised measures after spurious entrants and exits have been filtered out. In each linkage approach, entry and exit rates are slightly below 10 percent, which is in line with results for other European countries (Reynolds et al. 1994; OECD 2013). The traditional method, which captures more misclassified entrants and exits than the employee-flow method, yields entry and exits rates that most closely correspond to the benchmark results.

In contrast to firm turnover, average sizes and the employment shares of entrants and exits are considerably revised downwards after spurious entrants and exits have been filtered out. The lower panel of Table 3 shows the bias in the measures compared to the benchmark results. If based on the unedited data, average entry and exits sizes are overestimated by 64 percent and 78 percent respectively, and the job reallocation rates by 81 percent to 102 percent respectively. The employee-flow method strongly reduces these biases and produces results close to the ones that are obtained by both linkage methods combined. The traditional method only corrects the initial biases by about half. The large number of additional links that this method identifies in the smallest size class (see Table 2) account for only small shares of aggregate job reallocation and contribute little to bias reduction.

Figure 1 Employment distribution of entrants and exits



Note: Annual averages over the 2003-2012 period.

The employment distributions at entry and exit shed more light on these results. The left panel of Figure 1 presents the distribution of total employment created by new firms at entry, and the right panel shows the employment distribution of firms in the year of exit. The top lines represent results based on unedited data, while the other three lines show the results after implementation of the linkage procedures. Missing linkages strongly shift the distributions to the right. The unedited data falsely suggest that that an important amount of jobs is created by medium and large entrants, and likewise that more than half of job loss due to exit is brought about medium and large firms exiting the market. These patterns are only moderately corrected by the traditional method. The method fails to identify an important part of spurious entrants and exits in larger size classes, and leaves a considerable upward bias in the middle and right tale of the distributions. The employee-flow method, by contrast, strongly reduces the initial biases and reveals that job creation by entrants and job destruction by exits is highly concentrated in the smallest size classes. The results obtained by this method closely approximate the right-skewed distributions of the benchmark method.

Many studies have found that the firm size distribution at entry is highly right-skewed and that the likelihood of a firm's exit declines with its size (Cabral and Mata 2003; Caves 1998). Even a small number of larger entrants or exits may however represent important employment shares at entry or exit.

This is indeed suggested by the results based on unedited data and the traditional method. Improved longitudinal data, however, reveal that small firms do not only represent the major part of units but also the major part of employment at entry and exit. The benchmark results show that firms that start with less than 10 employees represent more than 80 percent of total employment at entry, and likewise that firm exits in these size classes account for 74 percent of job destruction due to exit. Entrants and exits with more than 50 employees barely contribute to job creation by entry and job destruction by exit.

The benchmark results are in sharp contrast with statistics reported by Eurostat for several European countries (Figure A.1 in the Appendix). The statistics are derived from harmonized national business registers that follow the Eurostat-OECD recommendations on record linking discussed above. The employment shares of larger entrants and exits differ widely by country and raise questions about the use of the traditional linkage methods for obtaining comparative results. Results for some countries are in line with our benchmark results, but other countries report high employment shares of larger entrants and exits. Similar large country differences are reported by OECD (2013).

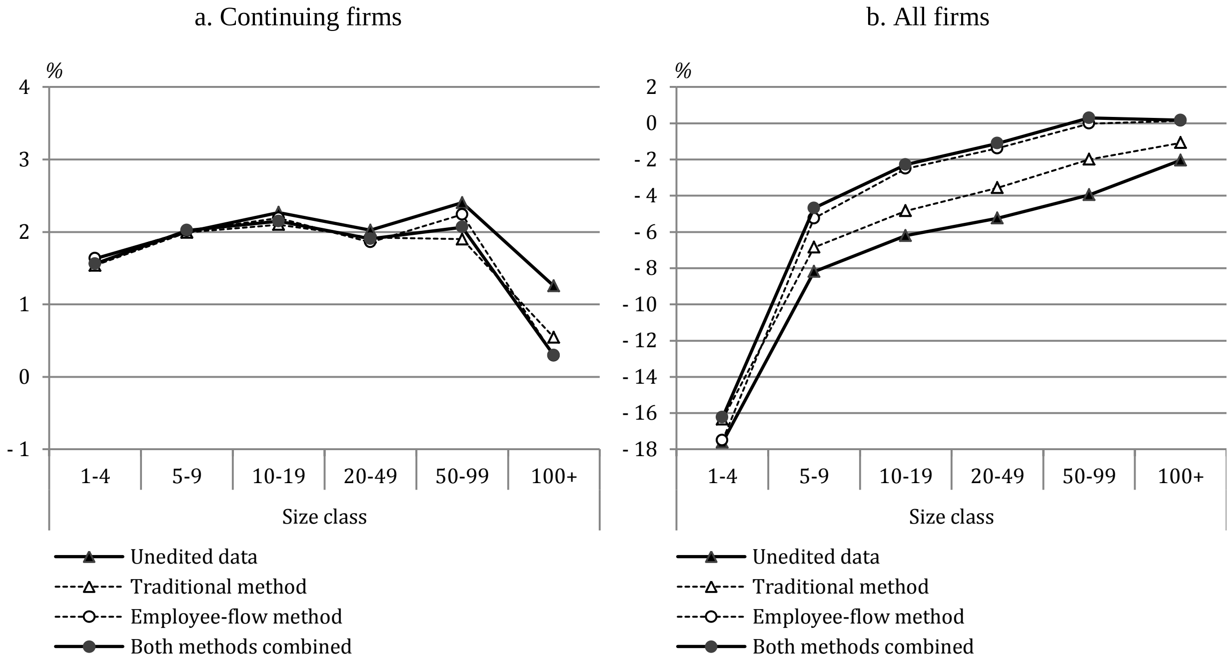
5.2 Size profile of firm growth

As explained in Section 3.3, firm-level growth rates are revised when exits are reclassified as continuing firms, or continuing firms are linked to another firm identifier. This section presents an example of the sensitivity of firm-level growth estimates to missing longitudinal linkages. Is shows estimations of the relationship between firm size and the firm growth rate, both in the sample of all firms, and in the sample of continuing firms only. The regression model and estimation method are described in the Appendix D.[13] The coefficient estimates with standard errors are reported in Table A.4 in the Appendix.

From the previous results, it can be expected that growth rates of large firms will be more sensitive to longitudinal linkage errors than the ones of small firms. Because missing linkages strongly affect the exit population in particular, it can also be expected that growth rate estimates of all firms will be affected more severely than those of continuing firms only.

[13] Regressions of continuing firms include 1.6 million firm-year observations; those of all firms 1.7 million.

Figure 2 Firm-level growth rates (in percent) of continuing firms and all firms



Note: Annual averages over the 2003-2012 period.

The left panel of Figure 2 plots the size coefficients of the regressions for firms that continue between *t-1* and *t*. The point estimates represent mean employment growth rates of a given size class of firms, which are the net result of job creation by expanding firms and job destruction by contracting firms. As expected, results for larger firms are more sensitive to missing links between identification numbers. The average growth rate of firms with over 100 employees is 1.3 percent per year when based on the unedited data, but revised downwards to 0.3 percent in the benchmark results. Revisions based on the traditional and employee-flow method are quite similar. Improved longitudinal linkages also increase the precision of the firm-level estimates, as can be seen from the reduction in the standard deviations presented in Table A.4.

The right panel of Figure 2 reports the growth rate estimates for all active firms in *t-1*, i.e. including the ones that exit in *t*. The plotted curves show a positive relationship between firm size and growth, which is mainly explained by the higher exit rates in smaller size classes. Growth rates in all size classes are revised upwards after the linkage procedures. This means that reclassifying spurious exits as continuing firms overrules the downward revision in growth rates of continuing firms. The revision is most substantial in size classes between 5 and 100 employees, where the benchmark results are 4 percentage points higher than the ones obtained with the unedited data. Replacing the -2.0 growth rates of spurious exits with values closer to the mean also yields a considerable gain in the precision of growth estimates: standard deviations given by the benchmark results are about one-fourth smaller than those based on the unedited data. The employee-flow method proves to be more effective for reducing bias in the growth rates in size classes above 5 employees: both the mean and standard deviations are close to the estimates based on the benchmark method. This results is again explained by the greater power of the method to capture spurious exits in medium and large size classes. The traditional method, which misclassifies an important share of these firms, reduces the biases by only half.

5.3 Job creation and destruction

The results up till now suggest that a poor strategy to identify real entrants and exits, and true employment gains and losses when firms change structure will strongly affect aggregate job reallocation rates. The reason is that misclassifications increase with firm size, and thus account for important shares of total job reallocation. To document aggregate measures of job creation and destruction, we follow Davis et al. (1996a), and decompose the net employment growth rate into the job creation rates by entry and by expansion, and the job destruction rates by exit and by contraction (see Appendix E).

5.3.1 Total job creation and destruction rates

Table 4 shows that missing linkages introduce a considerable upward bias in each of the four job reallocation rates. Average annual employment growth is 1.03 percent in the period of observation (2003-2012).[14] The benchmark results show that the net employment growth is the result of an average annual job creation rate of 7.06 percent and a job destruction rate of 6.03 percent. When based on the unedited data, these measures are overestimated by 28 and 32 percent, respectively. The biases are primarily due to the 81 percent overestimation of the job creation rate by entry and the 102 percent overestimation of the job destruction rate by exit. The relative biases in the job reallocation rates of expanding and contracting firms are much smaller, but further add to the overestimation of the total rates.

In line with the previous results, Table 4 shows that the employee-flow method is more effective for obtaining accurate job reallocation rates and yields job reallocation rates that are close to the benchmark results. As expected, the traditional method leaves a substantial upward bias in the measures. This is most noticeable in the 50 percent overestimation of the job creation rate by entry and the 48 percent overestimation of the job destruction rate by exit.

[14] By definition, net employment growth is not affected by the linkage procedures since they only reshape the reallocation of jobs across firms.

Table 4 Annual job creation and destruction rates

	Net growth rate	Job creation rate			Job destruction rate			Net growth rate established firms [1]
		Total	By entry	By expansion	Total	By exit	By contraction	
a. Rates by linkage method (%)								
Unedited data	1.03	9.01	2.52	6.48	7.98	3.06	4.92	-1.49
Traditional method	1.03	8.17	2.09	6.08	7.14	2.24	4.89	-1.05
Employee-flow method	1.03	7.24	1.49	5.75	6.21	1.64	4.57	-0.46
Both methods combined	1.03	7.06	1.39	5.67	6.03	1.52	4.51	-0.36
b. Percent bias vs. both methods combined								
Unedited data		*28%*	*81%*	*14%*	*32%*	*102%*	*9%*	
Traditional method		*16%*	*50%*	*7%*	*18%*	*48%*	*9%*	
Employee-flow method		*3%*	*7%*	*1%*	*3%*	*8%*	*1%*	

Note: Annual averages over the 2003-2012 period.
[1] Sum of job creation rate by expansion and job destruction rates by exit and contraction.

Improved longitudinal firm linkages not only reduce total job reallocation, but also considerably affects the contribution of different classes of firms to net employment growth. First, new firms contribute less to job creation, and established firms destroy less jobs than the traditional method suggest. The benchmark results show that jobs created by new firms represent a mere 1.39 percent of total employment in a given year (column 3), while the traditional link method overestimates the contribution of entrants to job creation by 50 percent. On the other hand, job destruction by incumbent firms (last column) proves to be much smaller if based on good quality longitudinal data. Net annual employment growth of established firms is not -1.05 percent, as suggested by the traditional method, but only -0.36 percent.

Secondly, applying improved linkages considerably changes the allocation of absolute employment growth across different firm size classes. This is illustrated in Figure A.2 in the Appendix, which presents net employment growth by firm size class, expressed in absolute numbers in the total Belgian private sector. The upward corrections in the firm-level growth estimates (Section 5.2) have a greater impact on absolute net job creation of larger firms, as they account for more substantial shares of employment in the economy. The negative contribution of large firms to employment growth reported by the unedited data and the traditional method even reverses into a positive one if longitudinal linkages are accurately identified. The improved measurement challenges the common perception that small firms are the engine of job creation. Instead the results suggest that large established firms contribute a great deal more to employment growth than smaller ones.

5.3.2 Annual variation in job creation and destruction

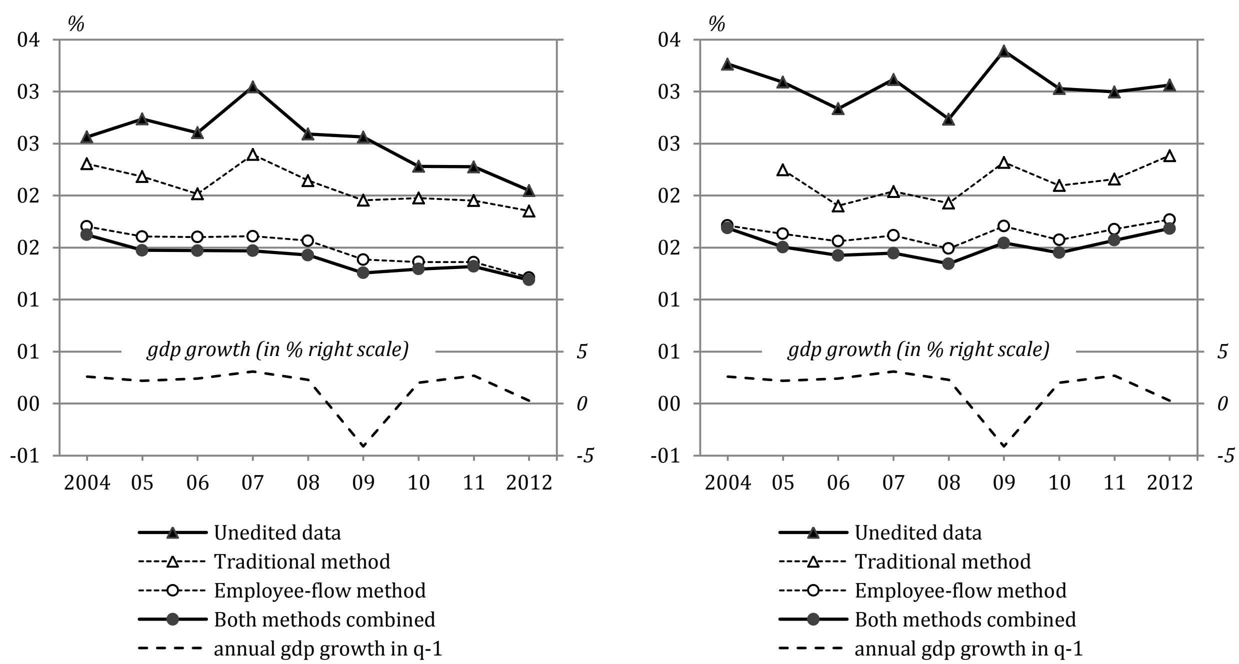
The bias in the job reallocation rates varies significantly over time. Fluctuations in the number of medium or large firms that change identification number of firm structure strongly affect annual measures. This is most clearly illustrated in the job reallocation rates by entry and exit.

Figure 3 shows that annual variation in the job creation rate by entry and the job destruction rate by exit is strongly reduced when spurious entrants and exits are filtered out. The unedited data and the traditional method report large annual fluctuations, while the employee-flow and benchmark results reveal that both job creation by entry and job destruction by exit are rather non-volatile, with the largest year-to-year change corresponding to the recession period of 2009. One typical pattern is explained by ID changes: fluctuations in the employment shares of firms that change identification number lead to symmetric increases and decreases in the job creation rate by entry and the job destruction rate by exit. This effect is most noticeable in the results for 2007: the unedited data report a peak in both the entry and the exit rates, which is entirely absent in the benchmark results.

Figure 3 Annual job reallocation by entry and exit

a. Job creation by entry

b. Job destruction by exit



Is this stable pattern of job reallocation from entry and exit more plausible than the annual fluctuations obtained using the traditional method? The correlation of the entry rate with GDP growth suggests it is. Business formation is considered to be procyclical, and especially job creation by entry is found to covary positively with output growth (Campbell 1988). The employee-flow results strongly reflect this feature; annual changes in the job creation rate by entry show a high positive correlation with GDP growth of the previous quarter (0.86), which adds support to the reliability of this identification strategy for entry and exit. The correlation is only half as large (0.42) for entry rates based on the traditional method. Misclassifying spurious entrants thus introduces substantial spurious variation in job creation by entry over time, which weakens the correlation with the business cycle. ID changes, mergers and split-ups are indeed mainly driven by legal, tax or administrative motivations and less by macro-economic fluctuations.

6 Conclusion

We have shown that even if an administrative register is used that includes good quality longitudinal firm linkages from the start, measures of employment dynamics are strongly biased. Large firms are disproportionally affected by the longitudinal linkage problem and introduce an upward bias in job creation and destruction rates, especially at entry and exit. Missing linkages lead to employment distributions at entry and exit that are biased towards larger firms, to spurious variation in annual job reallocation, and to an underestimation of firm-level growth rates in medium and large size classes. Firm turnover measures are only slightly overestimated.

The employee-flow method captures missing links in larger size classes well and most effectively reduces the initial biases. It produces empirical measures that are close to benchmark results obtained by using both linkage methods combined, while the traditional method reduces bias by only half. An optimal longitudinal research dataset is obviously obtained by using both traditional and employee-flow record linking. If only one method is used, the employee-flow method is clearly more preferable.

An additional advantage of the employee-flow method lies in international comparability. The method uses an economic definition of firm continuity, tracing one of the firm's key production factors, the stock of employees, to identify firms that operate continuously but change identification number or firm structure. This definition is translated in linkage algorithms that follow generally applicable rules and which are, unlike the traditional method, independent of country-specific data characteristics. Using employee-flow methods to harmonize longitudinal business databases for research could not only produce more reliable but also more comparable results across countries.

Appendix

The Appendix is available on request.

References

Abowd, J.M., & Kramarz, F. (1999). The Analysis of Labor Markets Using Matched Employer-Employee Data. In O.C. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics Volume 3, Part B* (pp. 2629–2710). Elsevier.

Acs Z. J., & Armington, C. (1998) Longitudinal Establishment and Enterprise Microdata (LEEM) Documentation. CES 98–9, Center for Economic Studies, US Bureau of the Census.

Acs, Z. J., & Audretsch, D. B. (1989). Births and Firm Size. *Southern Economic Journal*,56(2), 467-475.

Albaek, K., & Sorensen, B.E. (1998). Worker flows and job flows in Danish manufacturing, 1980-91. *The Economic Journal*, 108, 1750–1771.

Baldwin, J.R., Dupuy, R., & Penner, W. (1992). Development of longitudinal panel data from business registers: the Canadian Experience. *Statistical Journal of the United Nations*, 9, 289-303.

Baldwin, J.R., & Gorecki, P.K. (1987). Plant Creation versus Plant Acquisition. The Entry Process in Canadian Manufacturing. *International Journal of Industrial Organization*, 5(1), 27-41.

Bartelsman, E., Scarpetta, S., & Schivardi, F. (2005). Comparative analysis of firm demographics and survival: evidence from micro-level sources in OECD countries. *Industrial and Corporate Change*, 14(3), 365-391.

Benedetto, G., Haltiwanger, J.C., Lane, J., & McKinney, K. (2007). Using worker flows to measure firm dynamics. *Journal of Business and Economic Statistics*, 25(3), 299–313.

Cabral, L., & Mata, J. (2003). On the Evolution of the Firm Size Distribution: Facts and Theory. *The American Economic Review*, 93(4), 1075-1090.

Campbell, J.R. (1998). Entry, exit, embodied technology, and business cycles. *Review of Economic Dynamics*, 1(2), 371–408.

Caves, R.E. (1998). Industrial Organization and New Findings on the Turnover and Mobility of Firms. *Journal of Economic Literature*, 36(4), 1947-1982.

Caves, R.E., & Porter, M.E. (1977). From Entry Barriers to Mobility Barriers: Conjectural Decisions and Contrived Deterrence to New Competition. *Quarterly Journal of Economics*, 91(2), 241-262.

Clayton, R. L., & Spletzer, J.R. (2009). Business Employment Dynamics.In T. Dunne, J. Bradford Jensen, & M. J. Roberts (Eds.), Producer Dynamics: *New Evidence from Micro Data* (pp. 125-147). Chicago: University of Chicago Press.

Davis, S., Haltiwanger, J.C., & Schuh, S. (1996a). *Job creation and destruction*. Cambridge: MIT Press.

Davis, S., Haltiwanger, J.C., & Schuh, S. (1996b). Small business and job creation: Dissecting the myth and reassessing the facts. *Small Business Economics*, 8(4), 297–315.

Dunne, T., Roberts, M. J., & Samuelson, L. (1988). Patterns of Firm Entry and Exit in U.S.Manufacturing Industries. *RAND Journal of Economics*, 19(4), 495-515.

Dunne, T., Roberts, M. J., & Samuelson, L. (1989). The growth and failure of U.S. manufacturing plants. *Quarterly Journal of Economics*, 104(4), 671–698.

Eriksson, T., & Kuhn, J.M. (2006). Firm spin-offs in Denmark 1981–2000. Patterns of entry and exit. *International Journal of Industrial Organization*, 24(5), 1021–1040.

Eurostat-OECD (2007). Eurostat-OECD Manual on Business Demography Statistics. Luxembourg: European Communities and OECD.

Geroski, P.A. (1995). What do we know about entry? *International Journal of Industrial Organization*, 13(4), 421–440.

Geurts, K., & Van Biesebroeck, J. (2014). Job creation, firm creation, and *de novo* entry. CEPR Discussion Paper 10118.

Geurts, K., & Vets, P. (2013). Employee flows to study firm and employment dynamics. DynaM Working Paper, KU Leuven.

Haltiwanger, J., Jarmin, R.S., & Miranda, J. (2013). Who creates jobs? Small versus large versus young. *Review of Economics and Statistics*, 95(2), 347–361.

Hethey T., & Schmieder, J. (2013). Does the use of worker flows improve the analysis of establishment turnover? Evidence from German administrative data. NBER Working Paper 19730.

Jarmin, R.S., & Miranda, J. (2002). The Longitudinal Business Database. Center for Economic Studies Discussion Paper CES-WP-02-17.

Jovanovic, B. (1982). Selection and the evolution of industry. *Econometrica*, 50(3), 649–470.

Maliranta, M., & Nurmi, S. (2004). Analyzing entrepreneurship with the Finnish Linked Employer-Employee Data (FLEED), ETLA Discussion Paper 920, The Research Institute of the Finnish Economy.

Mata, J. (1993). Entry and type of entrant. Evidence from Portugal. *International Journal of Industrial Organization*, 11(1), 101-122.

Mata, J., Portugal, P., & Guimaraes, P. (1995). The survival of new plants: Start-up conditions and post-entry evolution. *International Journal of Industrial Organisation*, 13(4), 459–481.

OECD (2013). *Entrepreneurship at a Glance*. Paris: OECD Publishing.

Persson, H. (2004). The Survival and Growth of New Establishments in Sweden, 1987-1995. *Small Business Economics*, 23(5), 423-440.

Pinkston, J.C., & Spletzer, J.R. (2002). Annual measures of job creation and job destruction created from quarterly microdata. *BLS Statistical Survey Papers*. Washington: Bureau of Labor Statistics.

Reynolds, P., Storey, D.J., & Westhead, P. (1994). Cross-national Comparisons of the Variation in New Firm Formation Rates, *Regional Studies*, 28(4), 443-456.

Rollin, A.M. (2013). Developing a longitudinal structure for the national accounts longitudinal microdata file (NALMF). Proceedings of Statistics Canada Symposium 2013.

Storey, D.J. (1991). The birth of new firms—Does unemployment matter? A review of the evidence. *Small Business Economics*, 3(3), 167-178.

Wagner, J. (2007). Exports and Productivity: A Survey of the Evidence from Firm-level Data. *The World Economy*, 30(1), 60-82.