Schmollers Jahrbuch 133 (2013), 477 – 510
Duncker & Humblot, Berlin

Does the Use of Worker Flows Improve the Analysis of Establishment Turnover? Evidence from German Administrative Data

By Tanja Hethey-Maier and Johannes F. Schmieder*

Abstract

Administrative datasets provide an excellent source for detailed analysis of establishment entries and exits on a fine and disaggregate level. However, administrative datasets are not without problems: restructuring and relabeling of firms is often poorly measured and can create large biases. Information on worker flows between establishments can potentially alleviate these measurement issues, but it is typically hard to judge how well correction algorithms based on this methodology work. This paper evaluates the use of the worker flow methodology using a dataset from Germany, the Establishment History Panel. We first document the extent of misclassification that stems from relying solely on the first and last appearance of the establishment identifier (EID) to identify openings and closings: Only about 35 to 40 percent of new and disappearing EIDs with more than 3 employees are likely to correspond to real establishment entries and exits. We provide 3 pieces of evidence that using a classification system based on worker flows is superior to using EIDs only: First, establishment birth years generated using the worker flow methodology are much higher correlated with establishment birth years from an independent survey. Second, establishment entries and exits which are identified using the worker flow methodology move closely with the business cycle, while events which are identified as simple ID changes are not. Third, new establishment entries are small and show rapid growth, unlike new EIDs that correspond to ID changes.

Zusammenfassung

Administrative Prozessdaten, bieten die Möglichkeit, Betriebsein- und Austritte detailliert zu analysieren. Allerdings haben administrative Daten auch Probleme: Umstrukturierungen und Umbenennungen von Betrieben werden oft nicht richtig abgebildet und können größere Verzerrungen bewirken. Informationen über Arbeiterflüsse haben das Potential diese Messprobleme zu verbessern, aber es ist im Allgemeinen schwer zu beur-

* We would like to thank Stefan Bender, Till von Wachter, and workshop participants at the EESW (2009) and CAED (2010) for many helpful comments. We are grateful for support by the Institute for Employment Research in Nuremberg, and the Deutsche Forschungsgesellschaft (DFG). All remaining errors are our own.

teilen, wie gut Korrekturmechanismen auf Basis dieser Methode funktionieren. Diese Studie untersucht die Arbeiterflussmethode im Kontext eines Datensatzes aus Deutschland, dem Betriebshistorikpanel. Wir dokumentieren das Ausmaß von Missklassifikationen, die auftreten, wenn lediglich das erste und letzte Erscheinen von Betriebsnummern (EID) verwendet wird, um Neugründungen und Schließungen zu identifizieren: Lediglich zwischen 35 und 40 Prozent der neuen und verschwindenden EIDs mit mehr als drei Beschäftigten sind wahrscheinlich reale Betriebsein- und Austritte.Wir legen drei Beweisstücke vor, dass ein Klassifikationssystem basierend auf Arbeiterflüssen eine Verbesserung darstellt: Erstens, Betriebsgründungsjahre basierend auf unserer Methode sind deutlich höher mit den Gründungsjahren aus einer unabhängigen Umfrage korreliert. Zweitens, Ein- und Austritte, die über unsere Arbeiterflussmethode identifiziert werden, folgen dem Konjunkturzyklus sehr dicht, während Ereignisse, die als Änderungen der Betriebsnummern identifiziert werden, dies nicht tun. Drittens zeigen wir, dass identifizierte Betriebsneugründungen klein sind und schnell wachsen, was nicht der Fall bei einem Wechsel der Betriebsnummer ist.

1. Introduction

The availability of administrative firm and establishment level datasets has spurred new research in many areas of economics, spanning Labor, Industrial Organization and Trade. This work relies crucially on following establishments or firms over time. Unfortunately longitudinal firm/establishment identifiers (EID) often have problems: For example EIDs change spuriously due to changes of ownership or legal form, due to restructuring of the firm, or coding errors. When not taken into account a change of the EID will appear as a spurious establishment exit and a new entry. This is particularly problematic in research projects that focus on entering and exiting firms, for example analyzing the role of firm entries and exits for employment growth (and the lack thereof) during the great recession.[1]

A possible solution to this problem has gained popularity more recently (See for example Benedetto, Haltiwanger, Lane, McKinney 2007 and Vilhuber 2009): If individual workers can be linked to the EIDs and the individuals can also be followed over time, one can identify ID changes as events where large groups of worker simultaneously leave an exiting EID and enter a new EID. Thus using information on worker flows offers a way to generate consistent EIDs over time and correct for mistakes. While this method is quickly gaining popularity, it is not well known how big the potential bias is from using uncorrected establishment IDs or how well the method works.

In this paper we evaluate the method of using worker flows to identify true establishment entries and exits using German administrative data. We provide

[1] For examples, see Foster, Grim and Haltiwanger (2013), or Fort/Haltiwanger/Jarmin/Miranda (2012).

three separate ways to evaluate whether the worker flow method improves upon using simply uncorrected EIDs: First, we investigate whether EID entries and exits that we identify as corresponding to establishment entries and exits are more highly correlated with the business cycle, than EID entries and exits that we identify as spurious. Second, we combine the administrative data with an establishment level survey to compare establishment entry years derived using the worker flow method with the establishment entry year in the survey. And third, we analyze how establishment characteristics evolve around entries of EIDs. All three methods show that establishment deaths and births identified using the worker flow method, clearly correspond to real economic events and improve upon the simple measure of uncorrected establishment identifiers.

This paper is related to a number of papers that have documented problems with and attempted to correct longitudinal person identifiers. For example Abowd / Vilhuber (2005) describe the method used by the Longitudinal Employer-Household Dynamics Program (LEHD) at the U.S. Census and Vilhuber (2009) provides a broader overview. On the firm or establishment level, the problems are in some ways more difficult: while for person identifiers at least it is clear that the underlying unit of observation remains the same over time, firms and establishments change ownership, are restructured, break-up or relocate in ways that make it ambiguous what exactly the underlying unit of observation is that is to be tracked over time.

However a consensus has emerged that it is useful for economic research to distinguish cases where identifiers change due to a change in ownership, the legal form of the firm or simply a change of accountants. In this case the change of a firm identifier should not be counted as a firm exit in one and an entry in the next period. Furthermore it is generally thought that firm restructuring events such as merger, acquisitions and outsourcing should generally not be considered as components of job creation and destruction (For a discussion see Persson, 1999; Baldwin et al., 2002; Benedetto / Haltiwanger / Lane / McKinney, 2007; Vilhuber, 2009; Geurts / Ramioul / Vets, 2009; Abowd / Vilhuber, 2011).

To deal with problems of longitudinal linkages, researchers and statistical agencies have employed probabilistic matching methods based on similarities in partial firm identifiers as well as information about name, location and economic activity (Eurostat / OECD, 2007; Vilhuber, 2009). More recently information on worker flows between employers has been used, since it is usually presumed that if the work force is identical in two consecutive years, then there is a high probability that these records relate to the same firm or establishment. This approach has been used for administrative datasets, among others, in Italy (Revelli, 1996; Contini, 2007), Finland (Vartiainen, 2004), the U.S. (Benedetto et. al., 2007), and Belgium (Geurts et. al., 2009). This study follows most closely the approach taken by Benedetto et. al. (2007). Our main contribution relative to this literature is that rather than just using the worker flow methodology to correct for problems in the EIDs, we document how successful this approach is with fixing these identifiers.

As an illustration for the importance of the bias from not using corrected EIDs, we investigate the role of establishment turnover for job creation and destruction using corrected and uncorrected identifiers. The notion that producer entry and exit is an important form of reallocation of production factors and thus contributing to aggregate growth has inspired a long line of theoretical and empirical research. One aspect of this reallocation mechanism that has been particularly prominent in the political sphere is the role of this churning process in the creation and destruction of jobs. New and small producers are often referred to as an important job growth engine, while the demise of a plant is usually lamented for the number of jobs it destroys.[2] For this reason job creation and destruction has long been studied by economists to enhance the understanding of the business cycle and the adjustment processes in the economy (David / Haltiwanger / Schuh, 1996; Bartelsman / Scarpetta / Schivardi, 2005; Brown / Haltiwanger / Lane, 2008). These studies typically decompose net job creation into the contributions of entering and exiting firms in addition to reallocation between existing firms. We demonstrate that in the German administrative data, using only EID entries and exits may dramatically overstate, by as much as 100 percent, the role of establishment turnover for job creation and destruction. Correcting for spurious EID entries and exits reduces the absolute measures for job creation and destruction by up to 13 percent and aligns them closer with the business cycle.

This paper continues as follows: Section 2 discusses the data we are using and describes our methodology, in particular our system to classify appearances and disappearances of EIDs. Section 3 takes this classification system to the data and evaluates how well the worker flow method does in identifying true economic events. In Section 4 we discuss the robustness of our results to choosing different cutoffs for classifying entries and exits and provide some evdience in support of the chosen thresholds. Section 5 provides the application to job creation and destruction measures, by showing the bias that arises from using uncorrected EIDs. Section 6 concludes.

2. Data and the Worker Flow Method

2.1 Data

The establishment history panel (BHP) is created from German social security records. Employers are required to file a report for all employees who are

[2] The impact of job destruction due to plant closings on the displaced workers has also received a lot of attention in the literature, see for example Jacobson / Lalonde / Sullivan (1993) and von Wachter / Song / Manchester (2009).

employed during a year. This report contains information on the duration of employment, the total pay over that period and a number of demographic variables (such as education, nationality, gender, and age). The pay information is generally very accurate (since it determines the social security contributions) but top coded. There is also information on industry, occupation and work status (full-time, part-time, apprentice) available. Since employers and individuals are uniquely identified through establishment and person IDs, it is possible to construct complete job histories for individual workers, and to follow establishments over time. The data covers all employment subject to social security contributions, but excludes certain types of government employees and the self-employed.[3] Overall about 80 percent of the working population in Germany is in the dataset. While this is a large part of the population, it is clear that for the study of establishment turnover this dataset omits establishments that consist only of workers not covered by social security (mainly government agencies and institutions such as schools) or only of self-employed such as law firms who only have partners – though in most cases there would at least be some staff who is employed and liable to social security contributions.

Establishments are identified on the basis of establishment identification numbers (EID). Those numbers are allocated to each organizational unit in a specific region and industry consisting of at least one worker liable to social insurance.[4] The definition of an establishment in this data does not necessarily correspond exactly to a meaningful economic unit like a firm or a plant. An establishment may consist of one or more branches. As long as they all belong to the same industry and authority district (Kreis) they might all be covered under the same EID. Once an establishment is assigned an EID this number remains constant over time. This holds especially if the establishment moves to another region or is temporarily closed. The latter prevents classifying a re-opened establishment as a true entry. On the other hand, an ownership or industry change triggers the assignment of a new EID to an establishment, despite not being a truly new opening.

The BHP is created by collapsing social security records data on the establishment year level.[5] Only employment spells that cover June 30th are used so that for each establishment and year there is a record with information on characteristics and size of the employees on this date. The resulting data is a panel comprising the universe of German social security liable employment since the year

---

3 Also marginal part-time employment had been exempt from social security until 1999, so that up to this date it is not included in this data.

4 Since 1999 establishments with at least one marginal part-time worker are also assigned an EID.

5 The social security data is available to the scientific community in several different forms, ranging from individual level panel data (the IABS) to linked employer-employee data (the LIAB). For more information see http://fdz.iab.de.

1975. Our analysis is based on BHP data for the time period 1975 – 2004. The strength of this data is clearly its large scope (about 2 million observations per year covering about 25 million jobs) and time span. One important weakness, and the motivation for this paper, is that it is difficult to identify establishment entry and exit in the BHP. While for each EID it can be easily determined when it appears for the first and last time, it is not clear that these dates correspond to true entries and exits. An important concern is that if an EID changes for other reasons, this would appear as an exit and an entry without any corresponding economic event. That this can happen is acknowledged in the documentation of the BHP (Dundler et al., 2006), but it is hard to judge how often this actually happens and whether this biases empirical work that ignores the issue.

Establishments and Firms

It is helpful to clarify what we mean by establishment entry and exit before discussing how to identify these events. We understand an establishment to be a local economic unit consisting of workers and capital, and producing some sort of goods or services. Examples are a manufacturing plant, a restaurant, a local branch of a bank, or a gas station. This is different from the â firmâ as an economic unit, which may consist of several establishments, which may create new or destroy old establishments, and which may buy or sell them. It can be the case that a firm disappears but an establishment belonging to the firm continues to exist (e.g., after being taken over by a competitor) and vice versa.

It is not completely clear under which conditions one would consider an establishment in year t to be the same establishment in year $t + 1$. If all workers are still employed at the same location but possibly by a different owner or as part of a different company, one would probably consider this a continuing establishment that experienced an ownership change. On the other hand if only the location is the same and the new owner replaced all old workers with new ones, one would likely consider this a new establishment. In between these two extremes the distinction becomes fuzzy and in practice somewhat arbitrary definitions will have to be made. In addition to ID Changes, which allow following an establishment from one year to another, and clear creations or destructions of establishments, it is also possible for establishments to break up into several units or for several establishments to merge.

For this paper we completely ignore the capital aspect of establishments (for data reason) and focus on the employee side. We therefore define a new establishment to be an establishment where a new group of workers get together and start producing something, and we define a continuing establishment to be an establishment where a large part of the workforce has been employed together in the previous year. We will also take care to classify break ups and spin-offs appropriately. Since we do not have direct information on ownership structure or firm identities, it should be kept in mind that we are limited in that dimension.

2.2 The Worker Flow Method for Correcting Establishment IDs

In this paper we directly address the problem of spurious EID entries and exits by providing and evaluating a new way to identify establishment entry and exit based on worker flows. Having access to the underlying social security records of the BHP we observe directly how many workers move between each establishment pair between two consecutive years. We will call all workers who move from an establishment A to an establishment B, a cluster of workers. Such a cluster will represent an inflow in establishment B and an outflow in establishment A. Using the individual level social security data, we created a dataset on all worker flows, where a unit of observation is one clustered flow. Of all the clustered inflows to an EID, we call the largest one (most number of workers) in a given year the maximum clustered inflow (MCI). Similarly we define the largest flow of all the clustered outflows in a year the maximum clustered outflow (MCO).[6]

Our strategy to classify new EIDs into new establishments, Spin-Offs, and id changes is based on whether the workers in a new establishment all come from the same EID or not. In practice this is done by looking whether not more than a certain percentage of the current work force at an entering EID was employed together in the previous year. To check this it is sufficient to know the total number of workers currently employed, and the maximum clustered inflow to the EID. Similarly, in order to classify exiting EIDs it is enough to have information on the maximum clustered outflow. We therefore restrict our flow data to the MCI and MCO and merge those to each establishment year observation in the BHP.

Classifying Entering Establishment IDs

Not all new EIDs are also new establishments since an EID can change for a number of reasons. However it is true that the way EIDs are assigned in Germany implies that almost all new establishments will receive a new unique EID.[7] This allows us to focus on new EIDs to identify new establishments.

---

6 In addition to inflows from other establishments, there are also workers that were not employed in a social security liable job on June 30th of the previous year. In our flow data we cannot distinguish between whether these workers were unemployed at that time or worked in a job not covered by our data (self-employed, government or jobs below the earnings threshold for social security). The MCI (and similarly the MCO) is the maximum of all inflows from other establishments, so if no workers come from other establishments the MCI would be 0.

7 Except for the qualifications in the data section of how an establishment is defined in the BHP, there is only one qualification: If a business owner essentially shuts down his business for a number of years and then reopens it, she may use the same EID again even though this may reasonably referred to as a new establishment by our definition.

Based on the previous discussion a new EID can correspond to either a new establishment or a continuing establishment. A new establishment is an establishment where the workforce consists largely of workers that have newly come together to the production process (either be as a new firm or as part of an existing firm).

Continuing establishments correspond to the case where a large fraction of the workforce at the new establishment was employed together in the year before. We will call the EID where the largest cluster of workers has been employed together in the prior year the predecessor. If the workers at the new EID that were employed together in the year before also constituted most of the predecessor's employment, then the new EID and the predecessor correspond to very similar working arrangements and we will thus call them the same establishment, that underwent a change of the establishment identifier.

The other possibility for a continuing establishment is that a large fraction of the workers have been employed together in the previous year, but that they did not actually represent a large fraction of the workforce of the predecessor. We call this case a Spin-Off or break up, since a part of the predecessor is spun-off to create a new production unit. This can be further distinguished in whether or not the predecessor continues to exist or not. If not, we refer to a Spin-Off as pushed, since the group of workers is pushed out by the closing/disappearance of the larger unit. If the predecessor continues to exist we label the Spin-Off as pulled. While we use the label 'spin-off' here, it is important to note that without more context or information, this may encompass a number of different economic events. For example if an existing firm creates a new branch or plant (which for industry code or accounting reasons is assigned a new EID), this may be associated with a flow of workers from existing plants. Since such a new EID would still be part of the same company, calling this spin-off may be a bit misleading. Exploring the economic significance of these spin-off events is a particularly interesting avenue for future research. Also note that some new EIDs do not fit any of these patterns very well. We will come back to those later. From this discussion we can classify new EIDs into the following five broad categories:

- New establishments:
  A group of workers who come together to form a new production unit
- Continuing establishments: Spin-Off/Break Up pushed
- Continuing establishments: Spin-Off/Break Up pulled
- ID Change (because of ownership change, take over, change of legal form, restructuring)
- Other/Not classifiable/Unclear.

In order to apply these classifications to the data it is necessary to define cutoffs for what it means that most workers did not work together in the previous year etc. Our cutoffs and classification system follow Benedetto et. al. (2007) and are displayed in Table 1. For very small establishments the ratio of MCI to employment is not a very meaningful statistic (since for example for an establishment with exactly one worker in its first year this ratio can only be 0 or 1). We therefore put all establishments with less than 4 workers in the first year into an extra category which we call New Establishments (small). For the establishments with more than 3 employees we use the MCI to categorize them. If the MCI is less than 30 percent of all inflows in the first year of an EID, we call this a medium or big New Establishment (med & big). For 30 to 80 percent of MCI / inflows and less than 80 percent MCI / predecessor employment we put the new EID into a category which we call New Establishment (fuzzy) to indicate that these are likely new establishments but that there is some possibility of misclassification.

Most establishments with a higher than 80 percent MCI / inflow ratio can be considered to be continuing establishments. To distinguish between the different continuing establishment categories it is necessary to look at the predecessor. If the MCI corresponds to less than 80 percent of the predecessor's total employment (in the previous year), we call the continuing establishment a Spin-Off, if it is more than 80 percent and the predecessor exits we call it an ID-change. If the predecessor exits from the previous to the current year, we call the Spin-Off pushed, otherwise pulled. The remaining fields seem odd combinations for various reasons and are thus labeled Unclear (we come back to this in the results section).

Classifying Exiting Establishment IDs

Our method for classifying exiting establishments follows the same principle. All exiting establishments with less than 4 workers are classified as small establishment deaths, since for those the ratio of MCO to employment in the last year is not a meaningful statistic. All establishments where the ratio of MCO to employment in the year before the exit is less than 30 percent are classified as atomized deaths. Exiting establishment IDs where the MCO / last employment ratio is between 30 and 80 percent are classified as fuzzy deaths. It is certainly debatable what the best classification for this group is. One could both imagine that establishments of this kind are true exits, where a relatively large chunk of workers happens to end up at the same establishment, or some kind of spin-offs or takeovers that only take a relatively small fraction of workers. Since we think that any cutoff is ultimately arbitrary we put them in a separate category, which allows us later to see the importance of this group. For symmetry with the entry classification we label establishments with less than 80 percent MCO / outflow ratio and more than 80 percent MCO / successor employment ratio Spin-Offs (in this cased pushed, since the predecessor exits).

Table 1

Classifying Entering and Exiting Establishments by Clustered Worker Flows

Panel A: Entries								
		Predecessor exits MCI / Predecessor Employment			Predecessor continues MCI / Predecessor Employment			No predecessor MCI=0
		<30%	30–80%	>80%	<30%	30–80%	>80%	
	MCI / Inflows							
≤3 empl.	-	New Estab (small)	New Estab (small)	New Estab (small)	New Estab (small)	New Estab (small)	New Estab (small)	New Estab (small)
>3 empl.	<30%	New Estab (med & big)	New Estab (med & big)	New Estab (med & big)	New Estab (med & big)	New Estab (med & big)	New Estab (med & big)	New Estab (med & big)
	30–80%	New Estab (fuzzy)	New Estab (fuzzy)	Unclear	New Estab (fuzzy)	New Estab (fuzzy)	Unclear	
	>80%	Spin-off pushed	Spin-off pushed	ID Change	Spin-off pulled	Spin-off pulled	Unclear	
Panel B: Exits								
		Successor is entrant MCO / Successor Employment			Successor is existing estab. MCO / Successor Employment			No successor MCO=0
		<30%	30–80%	>80%	<30%	30–80%	>80%	
	MCO / Outflows							
≤3 empl.	-	Small Death	Small Death	Small Death	Small Death	Small Death	Small Death	Small Death
>3 empl.	<30%	Atomized Death	Atomized Death	Spin-off pushed	Atomized Death	Atomized Death	Atomized Death	Atomized Death
	30–80%	Fuzzy Death	Fuzzy Death	Spin-off pushed	Fuzzy Death	Fuzzy Death	Fuzzy Death	
	>80%	Unclear	Unclear	ID Change	Take-Over / Restruct.	Take-Over / Restruct.	Unclear	

Notes: MCI stands for Maximum Clustered Inflow: the size of the largest cluster of inflowing current workers. Inflows stands for all the total number of workers that arrived since the previous year at an EID, which for a new EID is the same as total current employment. MCO stands for Maximum Clustered Outflows: the size of the largest cluster of outflowing current workers. Outflows are all workers that leave the EID until the next year.

Exiting EIDs where a very large fraction – we take 80 percent as the cutoff – of workers stay together indicate that these are not true exits. If these worker go to a new EID in the following year and this group makes up most of the workers at the new establishment ID, then we take this as a strong indication that this is actually simply a change of the EID and we classify this as an ID change. If the workers enter an existing EID and make up less than 80 percent of the workforce at this EID, this may correspond to a takeover of the exiting establishment and we label this takeover/restructuring. The remaining categories are labelled unclear again.

2.3 Applying the Worker Flows Method to German Establishments

Table 2 Panel A shows the total number of establishments in each of our seven entry categories, pooling all establishment entries from 1976 to 2003. The vast majority (83 percent) of all new EIDs are New Establishments (small), with the two second largest groups being the other two New Establishment classes, accounting for 6 percent each. The other categories account for far fewer establishments: ID-changes for about 0.8 percent and Spin-Offs (pulled) and Spin-Offs (pushed) for 1.7 and 1.1 percent respectively. About 0.9 percent are classified as Unclear. While thus 95 percent of all new EIDs appear to be truly new establishments (excluding the fuzzy category), and Spin-Offs and ID-changes appear to be pretty rare, this masks the fact that most of these new establishments are very small. The table therefore also shows total employment in each of these establishment classes (in the year the EID appears). This changes the relative importance of these categories substantially. ID Changes and Unclear entries now account for nearly 10 percent of employees in new EIDs. Spin-Offs combined have about 3 million employees in their first year out of a total of 17 million in new EIDs. New establishments still account for most employees (about 73 percent), but the group of small establishments is now much less important (though still the largest) while the fuzzy and med & big groups account for 4 and 3 million employees each. Given the ambiguity of the fuzzy new establishment category, the group of unambiguous establishment entries is thus significantly reduced when either considering employment weighted number (accounting for only 50 percent of all employment) or when considering only EIDs with more than 3 employees (accounting for only 37 percent of all new EIDs).

The impression that spurious entries, due to the non-new establishment categories, are more important among large new EIDs, is confirmed further when we break up the entry classifications by employment size in the first year (See Appendix Table A-2). By definition New Establishments (small) only appear in the smallest size class. Among the larger establishments it is apparent that the two new establishment categories become relatively less important as em-

Table 2

The Distribution of Entering and Exiting Establishment IDs over Entry / Exit Classifications (1976 – 2003)

	# Establishments	Percent	# Workers	Percent
Panel A: Entering establishment IDs				
New estab (small)	3,950,679	83.10	4,990,187	29.76
New estab (med & big)	295,800	6.22	3,026,472	18.05
New estab (fuzzy)	291,163	6.12	3,996,527	23.83
Spin-Off Pulled	78,900	1.66	2,222,568	13.25
Spin-Off Pushed	53,609	1.13	883,627	5.27
ID change	38,881	0.82	711,358	4.24
Unclear	45,196	0.95	939,927	5.60
Total	4,754,228	100	16,770,666	100
Panel B: Exiting establishment IDs				
	# Establishments	Percent	# Workers	Percent
Small death	3,494,502	82.88	4,321,132	30.01
Atomized death	293,127	6.95	3,377,142	23.46
Fuzzy death	239,519	5.68	3,247,262	22.56
Spin-Off Pushed	86,451	2.05	1,628,907	11.31
Takeover	36,652	0.87	661,479	4.59
ID change	37,625	0.89	681,140	4.73
Unclear	28,267	0.67	479,912	3.33
Total	4,216,143	100	14,396,974	100

ployment increases. It is probably not surprising that there are few truly new establishments that start out very big and those that do would often be new establishments set up by large multi-establishment firms or some kind of outsourcing of parts of an establishment, both of which may show up us Spin-Offs (pulled).

The total number of establishments in each exit category is reported in Table 2 Panel B. The Small Deaths account for the vast majority of exits, with nearly 83 percent. Among the exiting EIDs with more than 3 employees, the Atomized and Fuzzy Death categories are clearly the largest with 290,000 and 240,000 establishments respectively. Establishment deaths that are associated with a Spin-Off occurring, are less frequent, with a total of 86,000 establishments. Exiting EIDs that probably do not correspond to an actual dissolution of the establishment – Takeovers and ID Changes – make up about 37,000 establishments each. Finally about 0.7 percent of all establishments are classified as Unclear. Again these raw numbers overstate the importance of the Small Death category for employment. The numbers on employment in each of the categories reveal that the Small Death category, while still the largest, only accounts for about 30 percent of employment in exiting EIDs. The other two death categories are relatively more important for employment, having a share of about 23 percent each. Finally takeovers, ID Changes and Unclear exits do represent a sizable fraction of the workforce in exiting EID, representing a combined total of about 13 percent. This is again reinvorced when when we break up the exit types by establishment size in the year prior to exit (Appendix Table A-3): Among the smaller size classes the atomized and fuzzy death classes clearly dominate, accounting for most of the exits. However, these categories become less important among the larger establishments, where ID Changes and Takeovers are relatively more important. It is particularly interesting that among large establishment exits with 100 workers or more, less than one in four exits fall into the atomized death category. This clearly highlights the importance of controlling for spurious exits in studies of job-displacement.

To summarize, while spurious entries and exits are less important among smaller establishments, they become significantly more important when establishments are employment weighted or similarly when looking at larger establishments. While this supports the notion that it is potentially important to apply the worker flow method to control for spurious entries and exits, based on this categorization alone it is unclear whether the worker flow method does in fact improve in identifying exits and entries that correspond to real economic events.

3. Evaluating the Worker Flow Method

In this section we evaluate the performance of the worker flow categorization of establishment entries and exits using three independent methods, that allow us to judge whether the categories of entries and exits correspond closer to the economic events that we want to capture.

3.1 Cross Validation with Survey Data

As a first way to gauge whether the worker flow method is able to distinguish between real establishment entries and spurious entries, we compare the entry years that are implied by our method with the years of the foundation of the establishment according to a survey based self-assessment of the establishments. If the categorization of entering EIDs in the administrative data is meaningful, then for categories capturing true entries, the correlation between survey based foundation year of an establishment with the administrative birth year (based on the first appearance of the EID), should be significantly higher than for the other categories.

We derive the year of establishment foundation from the Establishment Panel (EP), a large, representative survey of German establishments.[8] The Establishment Panel is a panel of establishments that are interviewed yearly starting in 1993. The size of the panel varies over time but in recent years about 15,000 establishments are interviewed every year. The Establishment Panel can be linked on the establishment level to the establishment identifiers in the Establishment History Panel. We define as the birth year in the BHP, the year in which the EID first appeared. In the EP the birth year is the answer to the question when the establishment was founded. Establishments are also asked whether the foundation was a) a new firm or branch, b) a new establishment, or c) neither of these.

Table 3 shows the correlation coefficients between the birth years in the BHP and EP for the three EP establishment categories and the seven BHP establishment entry categories based on worker flows. Column (1) shows the correlation between survey birth year and administrative birth year in the 'new firm or branch' category of the EP. The correlation is highest in the New establishment small and med & big categories, providing support for our classification of these establishments as true entries. The correlation is weaker for the other categories, though also not zero. It is interesting that the correlation is still about 0.49 for ID-Changes. This might be because ID-changes could be associated with ownership changes and that survey respondents interpret 'founding year of the establishment' to be the year when ownership changed.

Table 3

Correlation Coefficients between Birth Year in Administrative Data (BHP) and Survey Data (Establishment Panel BP)

	New firm or branch (according to survey)	New Establishment (according to survey)	Not a new firm or branch (according to survey)
ID Change	0.49	–0.08	0.47
SpinOff Pulled	0.49	0.73	0.47
SpinOff Pushed	0.61	0.57	0.58
New Estab. (small)	0.68	0.45	0.56
New Estab. (med & big)	0.82	0.88	0.53
New Estab. (fuzzy)	0.55	0.52	0.54
Reason Unclear	0.51	0.58	0.51

Note: The table shows correlation coefficients between establishment birth years in the administrative dataset (BHP) and the survey (BP). Each column shows a different new eststablishment category according to the survey (BP).

8 More information regarding the EP can be found at: http: // fdz.iab.de

Column (2) shows the same correlations for the 'new establishment' category in the survey, again the correlation between birth years is highest for the New estab (med & big) category, close to 0.9. Strikingly the correlation between birth years is negative for the ID changes in this column and pretty low for the New Establishment (fuzzy) category, supporting our suspicion that this category may contain many spurious entries. Finally Column (3) provides correlations for the other category, which are generally weaker, as would be expected given that in this case the foundation year in the survey data may be spurious.

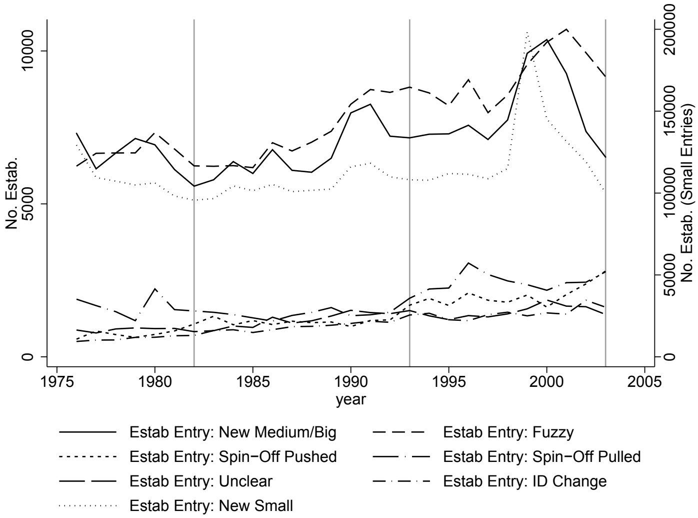
Overall the comparison with the survey based in formation supports our categorization in so far as that the correlations between birth years tend to be higher in the establishments categories that we would expect to correspond to true entries in both datasets. Notice that both datasets capture slightly different concepts. E.g. in the survey respondents may well state the age of the mother-firm rather than the establishment in a multi-level firm, thus measuring something different than the administrative establishment unit. Similarly respondents to the survey may or may not view an ownership change as a new 'foundation-year' of an establishment. So while the patterns seem to support our categorization, the imperfect correlations may not be surprising.

3.2 The Cyclicality of Entries and Exits

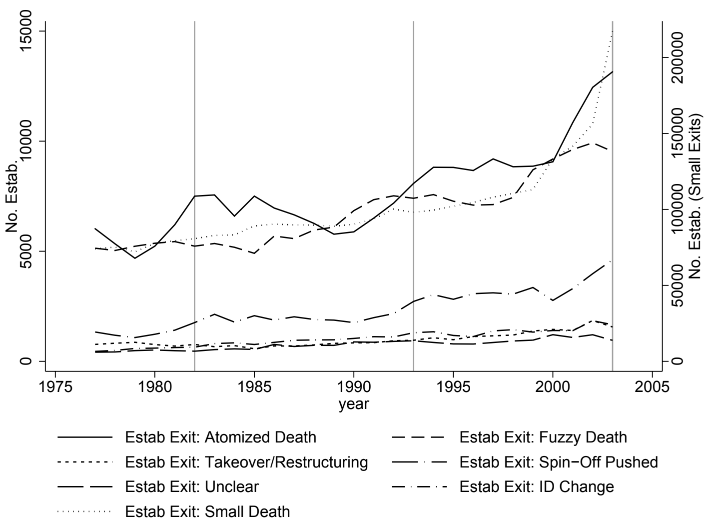
Economic upswings are usually associated with an increase in the formation of new establishments and firms, while recessions tend to be associated with plant closures and job destruction. If our entry and exit classification system does indeed capture differences in underlying economic events, then we should expect that the entry of EIDs classified as new establishments goes up in a boom and down in a recession, while the reverse should hold for EID exits classified as establishment exits. On the other hand the spurious entry and exit categories should be less correlated with the business cycle.

Figure 1 (a) shows the number of entering EIDs by entry category and year for West Germany.[9] On average there are about 120,000 new EIDs per year, with a slight increase to about 130 – 140,000 after 1990. 1999 (and to a lesser extent the following 2 years) is a clear outlier with a sharp spike in the New Establishment (small) category. In this year the reporting requirements for the

[9] For East Germany the data starts in 1991 and by focusing on 1992 and later we should not pick up establishments which are simply covered by the social security system for the first time. Nevertheless, as we show in the Web Appendix we still find a very large number of new establishments, more than 160,000, in 1992. Nearly all of them fall in the new establishment categories. The number of new EIDs drops sharply in 1993 and then shows a declining pattern, though with some outliers, across all categories until 2004.



(a) Entry Categories



(b) Exit Categories

Notes: The top figure shows the number of establishments in each of the 7 entry categories by year. Vertical lines indicate recession years. Data: Establishment History Panel. The bottom figure shows the number of establishments in each of the 7 exit categories by year. Vertical lines indicate recession years. Data: Establishment History Panel.

Figure 1: Number of New Establishments
in each Entry and Exit Category from 1976–2004

social security system were changed to cover marginally employed workers. While we attempted to correct for this by dropping these employment relationships, the underlying structure of the reporting rules make it impossible to correct for this perfectly which almost certainly explains the spike.[10]

The corresponding establishment exits are shown in Figure 1 (b). Again there appears to be an overall trend towards more establishment turnover throughout the 90s and early 2000s. The various true exit categories seem to increase similarly as the spurious exits, such as ID Changes.

In Figures 1 (a) and (b), recessions (1982, 1993 and 2003) are indicated by vertical bars. These figures give a visual impression of the cyclicality (and acyclicality) of the different time series: it appears that establishment entries in the medium / big category and small category are markedly lower in recessions, while the corresponding establishment exit categories increase in downturns.

We assess this more carefully by computing correlation coefficients between the time series of the different entry and exit categories and business cycle indicators. As business cycle indicators we use the growth rate of real GDP, as well as the year to year change in the unemployment rate measured in percentage points. Table 4 Panel A displays the correlation between number of establishments and number of employees in each of the seven entry categories with the two business cycle indicators. Since the change in the unemployment rate and GDP growth are quite highly negatively correlated (as one might expect from Okun's law), the patterns emerging from the two measures are pretty similar. Since several categories show strong increases over time, the raw correlation between such categories and the business cycle indicators (which are essentially trendless) will be highly affected by the long term trends and is thus not very informative. For this reason we detrend the establishment and employment time series using the Hodrick-Prescott filter.[11]

ID Changes and Spin-Offs Pulled are not strongly correlated with the business cycle and only the detrended time series show a weak (and statistically insignificant) counter cyclical correlation. For the Spin-Off Pushed category the correlation is very strongly counter-cyclical once the long term trend is

10 Apart from this outlier the number of EIDs in the New Establishment (small) category shows essentially no time trend. This is markedly different from all other categories which show fairly strong increases over time. Perhaps most striking is the fact that ID-Changes are more than three times as common towards the end of our sample period compared to the beginning. Similarly there is a very strong increase of both Spin-Off categories. There is also a pronounced increase in the Unclear and Fuzzy New Establishment categories, while the New Establishment (med & big) category shows only a moderate increase over time which reverts back to its starting value in the last 2 years.

11 We use a smoothing parameter value of 1600, which is commonly used for quarterly data, since we found that the more standard values for annual data take out too much of the cyclical variation. The results are very similar if instead of HP filtering, we simply take out a linear time trend.

Table 4

The Correlation Between Establishment Entry and Exit Categories and Business Cycle Indicators

	# Establishments		# Employees	
	Change in UR	GDP Growth	Change in UR	GDP Growth
Panel B: Entry Variables Detrended (Hodrick-Prescott Filtered)				
ID Change	0.28	–0.037	0.17	0.087
	[0.17]	[0.85]	[0.41]	[0.66]
Spin-Off Pulled	0.34	–0.31	0.22	–0.27
	[0.087]	[0.10]	[0.27]	[0.17]
Spin-Off Pushed	0.70*	–0.39*	0.48*	–0.31
	[0.000063]	[0.039]	[0.013]	[0.10]
New Small	–0.45*	0.38*	–0.64*	0.41*
	[0.021]	[0.043]	[0.00040]	[0.031]
New Medium/Big	–0.63*	0.48*	–0.69*	0.54*
	[0.00062]	[0.0096]	[0.000082]	[0.0028]
Fuzzy	–0.28	0.27	–0.31	0.35
	[0.16]	[0.16]	[0.12]	[0.064]
Unclear	–0.55*	0.45*	–0.12	0.19
	[0.0036]	[0.016]	[0.54]	[0.33]
Panel B: Exit Variables Detrended (Hodrick-Prescott Filtered)				
ID Change	0.25	–0.021	0.24	0.053
	[0.23]	[0.92]	[0.23]	[0.79]
Takeover/Restructuring	–0.016	0.26	–0.0090	0.23
	[0.94]	[0.19]	[0.97]	[0.24]
Spin-Off Pushed	0.70*	–0.37	0.66*	–0.33
	[0.000072]	[0.056]	[0.00022]	[0.091]
Small Death	0.13	0.15	0.31	–0.00099
	[0.52]	[0.46]	[0.12]	[1.00]
Atomized Death	0.68*	–0.34	0.65*	–0.32
	[0.00012]	[0.084]	[0.00029]	[0.11]
Fuzzy Death	–0.14	0.39*	–0.072	0.35
	[0.48]	[0.046]	[0.73]	[0.074]
Unclear	–0.39*	0.37	–0.014	0.23
	[0.048]	[0.058]	[0.94]	[0.25]

Note: The table reports correlation coefficients between the respective variables. The first two columns show the correlation between the number of establishments in each of the establishment categories with the business cycle indicators (in the column headings), the second two columns the correlation between the number of employees in the categories with the business cycle indicators. P-Values are given in brackets.

* indicates that the correlation coefficient is statistically significant on the 5 percent level.

taken out (correlation of 0.7 with the change in the UR). Since we think of these as spin-off which are forced by plant closings it makes sense that these are more common during downturns. On the other hand the New Establishment (med & big) and New Establishment (small) time series appear to follow the business cycle quite closely, showing clear and statistically significant correlations of around 0.4 to 0.6 with the business cycle measures.

The fact that only those entry categories which we consider to be relatively unambiguously new establishments are strongly procyclical indicates that our classification corresponds to real economically different events and we find this reassuring. Furthermore the ambiguity of the Unclear and New Establishment (fuzzy) categories is reflected in the weaker correlation with the business cycle, which points towards our suspicion that these categories correspond to true establishment entries as well as spin-offs and restructuring events.

For the exits in Table 4 Panel B, Atomized Deaths and Spin-Offs Pushed (which we argued should also be considered true exits) show nearly the same pattern of a very robust positive correlation with the change in the unemployment rate (about 0.7) and a weaker negative correlation with GDP growth. Interestingly the Small Death category is nearly uncorrelated with the business cycle, and thus shows a markedly different pattern than the New Small category. Also quite different from the respective entry categories, both the Fuzzy Death and the Unclear categories appear to be somewhat procyclical (although only marginally statistically significant), which may indicate that there are relatively few true exits in these categories and instead that they involve a significant amount of restructuring. The Takeover/Restructuring category is nearly acyclical as well as the ID Change category, which exhibits the same pattern as the corresponding entry category.

This evidence supports the common practice in displacement studies of the Jacobson, Lalonde and Sullivan (1993) type to view only worker separations from disappearing establishments that are large and atomized (in our parleance) as true displacements. The strong correlation of these establishment exits (in particular compared to the other EID exit categories) with economic downturns clearly indicates that these are more likely to correspond to real establishment closings and lumping all EID exits together may severely downward bias our displacement effects in the absence of such corrections.

3.3 The Evolution of Establishment Characteristics around Establishment Events

We now turn to how characteristics of establishments evolve around establishment events. We investigate the evolution of new EIDs over time depending on their entry type. There are two simple descriptive ways to achieve this. On the one hand one can pick a cohort of entering EIDs and follow them over time. On the other hand one can pick a year and analyze establishments of different ages in that year. The former approach has the problem that the variation with age is confounded by overall time trends, while the latter has the disadvantage that age is possibly confounded by differences of establishments across cohorts. We show results based on the latter approach.

Table 5 shows characteristics in establishments of different age in 2000 by their entry category. The first Panel shows how employment varies with establishment age (we speak of establishment age here even though we really mean the age of the EID, i.e., time since the first appearance of the EID). New establishments small and med & big show fairly strong employment growth over the first few years. For example New Establishments (small) have on average only 1 employee in their first year, but nearly 4 in their fourth year. New establishments (med & big) start out larger with about 12 employees, but this also quickly increases to 19 by age 4 and continues to rise afterwards. Since we would probably expect new establishments to grow this provides some support for our definition of new establishments. The New Establishments (fuzzy) category also shows increasing employment with age, but the relationship is not quite monotone. Also consistent with the fact that the other categories do not represent true establishment entries, they show no clear relationship between time since entry and employment.

The correlation between employment and establishment age can of course be driven by selection. This possibility is particularly important since new establishments have a very high probability of exiting again, so that the increase in average employment may be a simple composition effect. For this reason Table 5 Panel B shows how employment growth varies with establishment age. Here growth is computed on the establishment level (Employment current year minus employment last year divided by employment last year) and then averaged over the establishments. It is clear that the increase in employment in Panel A is not just driven by selection and instead all three new establishment categories show strong growth over the first couple of years, while the other categories show little growth.[12]

4. Sensitivity: What are the Correct Cutoffs for Defining Entries and Exits?

There is necessarily some arbitrariness in picking cutoffs for classifying establishment entries and exits. There is an inherent trade-off similar to the standard econometric trade-off between bias and efficiency. If the cutoff for the ratio of the maximum concentrated inflow to employment is very low, then the new establishment definition is most likely to correspond to only true entries. However in this case the number of observations is also quite low and a poten-

12 Schmieder (2013) investigates how these high growth rates in new establishments are associated with wages, showing a negative relationship between establishment age and wages within establishments.

Table 5

Establishment Size by Entry Category and Establishment Age

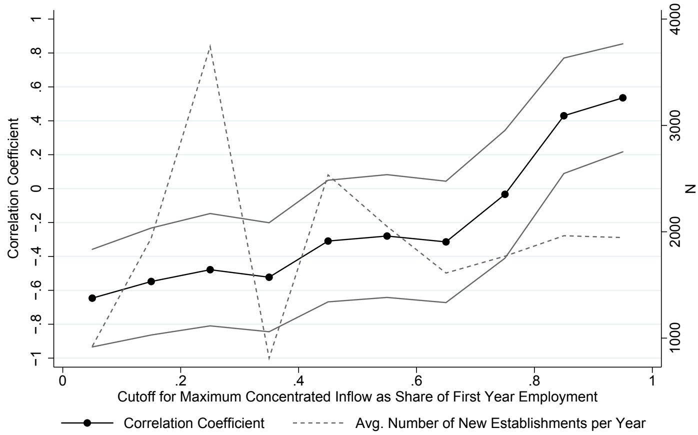
	Establishment Age in Years									
	0	1	2	3	4	5	6–10	11–15	16–20	21–25
Panel A: Establishment Size										
ID Change	19.1	18.8	18.5	21.8	18.9	16.2	13.9	17.3	17.8	15.9
Spin-off/pulled	36.6	45.1	41.4	42.0	45.3	48.3	40.8	31.7	32.7	36.8
Spin-off/pushed	20.0	21.6	14.3	16.0	16.9	18.1	14.3	16.6	18.0	16.7
New estab. (small)	1.1	1.6	3.1	4.0	3.7	4.4	4.3	4.9	5.6	6.2
New estab. (med & big)	11.7	14.6	17.1	18.4	19.1	20.4	20.1	23.1	23.4	24.9
New estab. (fuzzy)	15.2	18.8	18.7	19.3	21.9	20.2	18.7	20.6	19.9	24.4
Reason Unclear	23.1	26.0	21.2	24.3	27.4	25.0	19.2	21.7	21.1	27.5
Panel B: Employment Growth										
ID Change		–0.034	0.008	–0.001	0.025	–0.001	0.006	0.002	–0.001	–0.011
Spin-off/pulled		0.063	0.005	0.013	0.044	0.004	0.007	0.006	0.002	0.002
Spin-off/pushed		0.009	–0.001	–0.014	0.010	–0.005	0.004	–0.013	0.000	0.008
New estab. (small)		0.252	0.122	0.100	0.086	0.080	0.059	0.040	0.033	0.027
New estab. (med & big)		0.075	0.045	0.049	0.030	0.049	0.027	0.018	0.011	0.020
New estab. (fuzzy)		0.060	0.033	0.017	0.043	0.030	0.018	0.010	0.011	0.006
Reason Unclear		–0.031	0.026	0.000	0.027	0.024	0.008	0.005	0.011	0.000

tially large number of true new establishments are not captured by the definition. If the cutoff is increased, the definition will likely cover a larger share of real establishment entries, but will also include more 'false positives'. While the right definition depends on the research question at hand, in this section we present some evidence that the chosen cutoffs for establishment entries (MCI/inflows < 30%) and exits (MCO/outflows < 30%) may infact represent a reasonable choice for many applications.

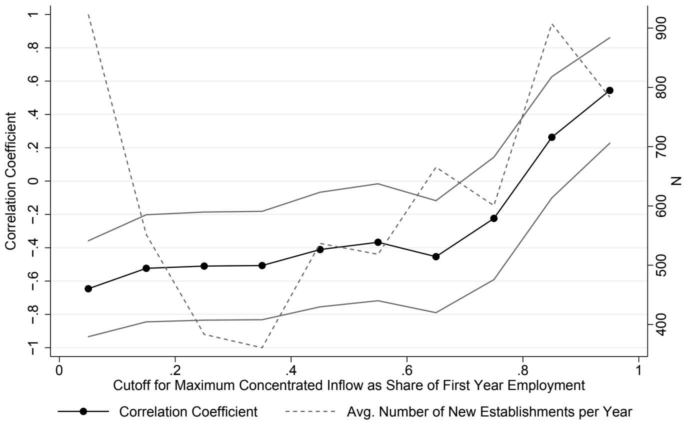
In order to investigate the role of different inflow cutoffs for defining establishment entries, we divided all new EIDs into bins of the ratio of the MCI and the employment in the first year. We use 10 bins for MCI/inflow ratios between 0 to 1, where the bandwidth of each bin is 0.1. We then compute the number of EID entries per year in each of those bins and compute the correlation coefficient between the number of EIDs and the change in the unemployment rate in the same way as we did for Table 4.

Figure 2 shows the correlation coefficients for the 10 bins (along with confidence intervals), as well as the average number of new EIDs per year in each of the bins. The top figure uses all new EIDs with more than 3 employees in the first year, while the bottom restricts the sample to EIDs that start out with at least 10 employees. Both figures show that the correlation between entries and the change in the unemployment rate is clearly negative (i.e., entries increase in good labor markets) for small cutoffs, as we would expect. As the cutoffs become larger the negative correlation becomes insignificant and eventually is reversed for the last two bins (MCI/inflows > 80%), providing strong support that EID entries with high MCI/inflow ratios are unlikely to be new establishments.

Turning to the number of establishments in each bin, the top figure is somewhat jagged with a spike in the 3rd bin. This is due to the fact that inflows are a



(a) Establishments with at least 4 employees in entry year



(b) Establishments with at least 10 employees in entry year

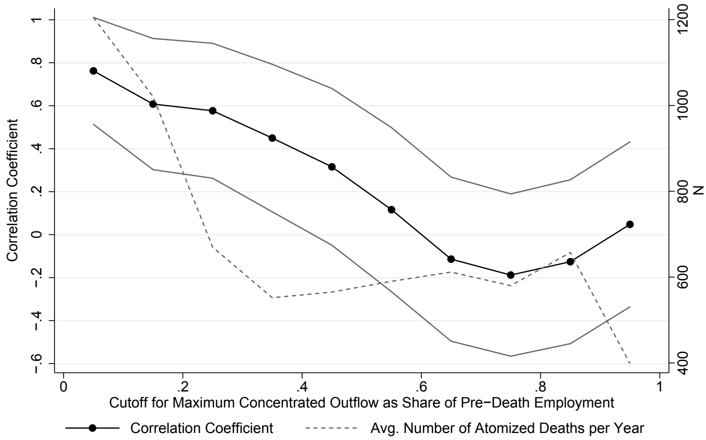
Notes: The figures show how the correlation between the number of new establishments per year and the change in the unemployment rate is affected by different thresholds for what constitutes an establishment entry. For each point, all new establishment where the share of the maximum concentrated inflow of entry year employment falls into the given cutoff range (e.g., [0 – 0.1] for the first bin) are calculated by year. Then the correlation between the (Hodrick-Prescott filtered) time series of establishment births and the year to year change in the unemployment rate is calculated. The left axis shows the correlation coefficient, while the right axis shows the average number of establishment births per year in the cutoff bin. The bands are 95% confidence intervals.

Figure 2: Correlation
between Number of New Establishments per Year
and the Business Cycle as a Function
of the Concentration of Employee Inflows

discrete number of employees and most EID entries are very small. E.g., establishments where all workers come from different EIDs in the prior year, still have an MCI/inflow ratio of 0.2 or 0.25 if employment in the first year is only 5 or 4 employees respectively. This explains why there are so few establishments in the first bin, since only EIDs with at least 10 employees could even show up in this bin. In the bottom figure, where we only look at EIDs with at least 10 employees in the first year, we do not have this issue due to the integer nature of inflows. Here we have a clear U-shaped pattern, with many new EIDs in the first bin, as well as for high MCI/inflow ratios. This figure strongly suggests that new EIDs are indeed divided into a large number of truly new establishments, but also a significant number of ID changes in the higher bins. Note that cutoff of MCI/inflows < 30% can be viewed as a sweet-spot, since it is roughly the turning point when the number of EIDs increases again and when the correlation coefficient is starting to not be significant anymore.

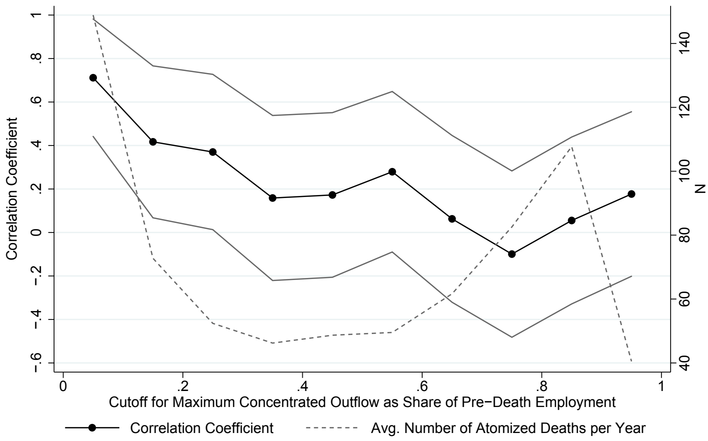
Figure 3 represents a parallel analysis for exiting EIDs. Since exits tend to be larger than entries in any case and many papers are interested in closings of larger plants, we show the results for EIDs with at least 10 (top figure) and 50 (bottom figure) employees.[13] The results are similar across both figures. The correlation between establishment exits and the change in the unemployment rate is, as expected, positive and highly statistically significant, for the first 3–4 bins but falls rapidly thereafter indicating that indeed only the lower bins correspond to true establishment deaths. Furthermore there is again a low point in the number of EIDs in the 3rd and 4th bin. Overall this indicates that the cutoff for defining atomized deaths, or true establishment exits, of MCO/outflows < 30% is a reasonable trade-off, providing a large number of observations that likely correspond to true exits.

We also conducted a careful sensitivity analysis of our other results where we varied the thresholds for inflows and outflows by values of 0.1 around the main threshold (e.g., defining new establishments to be EIDs where MCI/inflows < 20% or MCI/inflows < 40%). As one might expect from the low mass of EIDs in these ranges in Figures 2 and 3, this has only a relatively small impact on the number of observations in each entry and exit classification. Furthermore it has almost no impact on the correlations in Tables 3 to 5. This is again not surprising, given the results in Figures 2 and 3, since there the correlations are quite stable around smaller changes of the inflow and outflow cutoffs.



(a) Establishments with at least 10 employees in year prior to exit

[13] See Web Appendix for samples with EIDs with at least 4 and 30 employees. While the employee threshold of 4 and larger again displays the jaggedness in the number of EIDs per year due to the discrete nature of outflows, the 30 employee sample looks very similar to Figure 3.



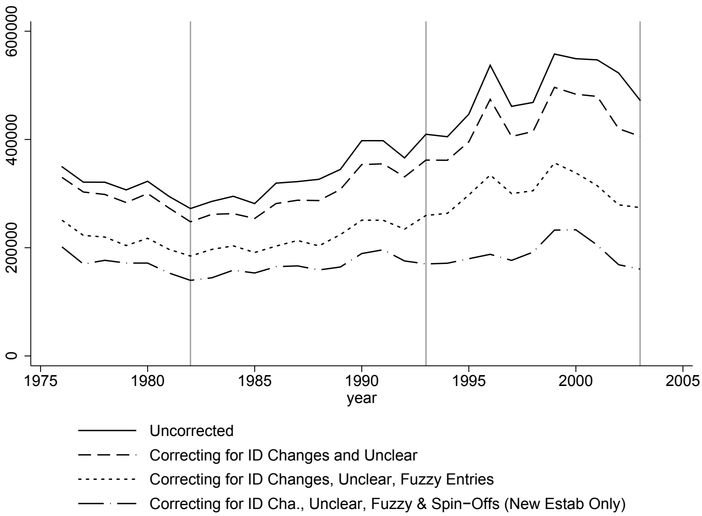
(b) Establishments with at least 50 employees in year prior to exit

Notes: The figures show how the correlation between the number of establishment deaths per year and the change in the unemployment rate is affected by different thresholds for what constitutes an establishment death. For each point, all establishment deaths where the share of the maximum concentrated outflow of pre-exit employment falls into the given cutoff range (e.g., [0 – 0.1] for the first bin) are calculated by year. Then the correlation between the (Hodrick-Prescott filtered) time series of establishment deaths and the year to year change in the unemployment rate is calculated. The left axis shows the correlation coefficient, while the right axis shows the average number of establishment deaths per year in the cutoff bin. The bands are 95% confidence intervals.

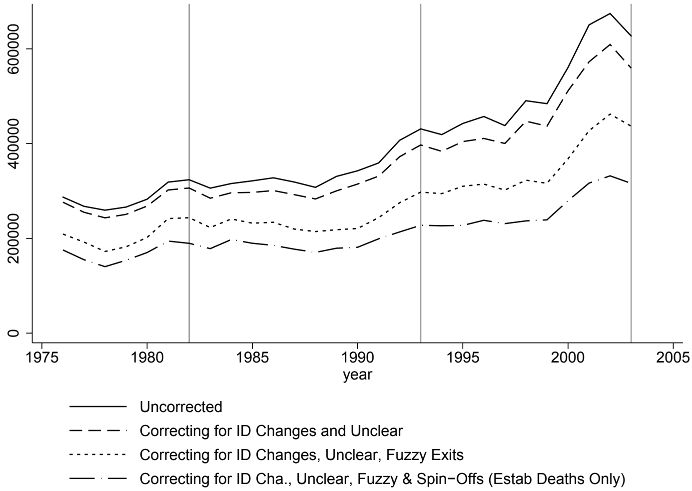
Figure 3: Correlation between
Number of Establishment Deaths per Year and the Business Cycle
as a Function of the Concentration of Employee Outflows

5. Application: Correcting Measures of Job Creation and Job Destruction

New establishments are often considered to be important contributors to overall job growth. However, as discussed before, spurious entries and ID Changes can significantly overstate the contribution by new entries. In order to assess the magnitude of this problem Figure 4 (a) shows job creation over time by new EIDs. The solid black line represents the uncorrected measure which corresponds simply to total employment in new EIDs in their first year of appearance. In a typical year, there are about 300,000 – 400,000 jobs in new EIDs, which represents about 25 percent of total job creation in the economy, or about 2 percent of all jobs. It is not completely clear, which of the entry categories should be considered new entries, or corresponding to true job creation. If we apply the most conservative measure and use only the New Small and New



(a) Entry Categories



(b) Exit Categories

Notes: The top figure shows corrected and uncorrected measures of job creation by year. Vertical lines indicate recession years. Data: Establishment History Panel. The bottom figure shows corrected and uncorrected measures of job destruction by year. Vertical lines indicate recession years. Data: Establishment History Panel.

Figure 4: Correcting Measures of Job Creation and Job Desctruction by New and Exiting Establishments for Spurious Entries and Exits

(med & big) category, the job creation number by new establishments is nearly cut in half and new establishments account for only about 13 percent of overall job creation. Furthermore the strong increase over time disappears and job creation by new establishments appears quite stable (though procyclical) in the long run. The figure also shows corrected measures which are less conservative and for example include the Fuzzy entries and Spin-Offs.

Figure 4 (b) shows the same for job destruction. Again the most conservative correction measure, shows a much smaller contribution of establishment exits to overall job destruction (about 15 rather than 25 percent) and decreases the long term time trend, although there is still a significant increase over time. Unsurprisingly our corrected measures for job creation and job destruction by entries and exits are also closer correlated with the business cycle.

We also analyzed total job creation and job destruction after taking out the creation and destruction by spurious entries and exits (Correcting in the same way as in Figures 4 (a) and (b)). Using the most conservative measure, it appears that the uncorrected overall job creation measure is about 13 percent higher (increasing in recent years) than our corrected measure, a quite significant upward bias.[14] Similarly the uncorrected total job destruction time series is about 11 percent higher than the corrected one. For less conservative corrections, the difference is smaller but still appears to be economically significant.

The impact of these corrections is strongest when we consider net job creation (defined as job creation minus job destruction) and destruction measures uncorrected and corrected for spurious entries and exits. In absolute numbers, the correction for net-job creation measures has a smaller impact, since the biases tend to cancel each other out. However since net job creation has a lower level (on average around 0), the relative bias (the ratio of spurious net-job creation to total net-job creation) in any particular year can be large and ranges from –60 to +30 percent in years where net job creation is close to zero. On average (over all years) the relative bias is about 16 percent, a quite significant number in economic terms.

6. Conclusion

Every year there is a large number of newly appearing and disappearing establishment identifiers in the data. In this paper we provide a way of classifying these events in order to distinguish true establishment entries and exits from ID changes and restructuring events. We find that clear cut establishment entries and exits account only for roughly half of the employment in entering and exiting EIDs. There is a large number of establishments which come out of Spin-

[14] See web appendix for detailed figures documenting this.

Off events or some sort of firm restructuring. There is also a sizable number of establishment identifiers, which disappear or appear in ways which are not easily classified. Finally there are sizable numbers of pure ID changes, particularly important among larger establishments.

Our rules to identify true entries and exits create time series that closely line up with the business cycle, while the other categories appear relatively acyclical. Across the board there are interesting time patterns which warrant further investigation. For example there has been a strong increase in establishment restructuring events in West Germany, while East Germany experienced a decline over the same time period.

Correcting job creation and destruction measures for spurious ID Changes and Restructuring events has very sizable effects on the overall numbers. Not correcting for such events overestimates the contribution of entries and exits to job creation and destruction by a factor of around 2. Furthermore overall job creation and destruction rates are severely biased and about 5 percent (for moderate corrections) to 10–13 percent (for more conservative corrections) lower when correcting for spurious events.

The bias created by time inconsistent establishment identifiers and firm restructuring events appears to be quite significant and may be even more problematic within particular industries, regions, or establishment size classes. It is hard to know exactly how big this problem is for the interpretation of previous studies which identified establishment turnover solely using the EID entries and exits (sometimes in conjunction with arbitrary size cutoffs), but it seems important to take the potential biases into account.

Fortunately our study indicates that using worker flows will allow for significant improvements of the firm linkages and thus improve the overall data quality of the BHP. Working together with the Research Data Center of the IAB, we have made the 6 crucial variables, on which all our definitions are based, available to users of the BHP, thus allowing researchers to either replicate our entry and exit categories, or create their own classification system. Several papers have already made use of our approach to study entries and exits. For example Fackler, Schnabel and Wagner (2012) study the role of establishment size and age for establishment exit patterns using the BHP in combination with the extension files containing the flow variables created by us. In addition to classifying entries and exits, these variables should also be useful for other purposes. For example Schmieder, von Wachter and Bender (2010) use the same information on worker flows to distinguish true Mass-Layoffs from spurious exits and spin-off events to study earnings losses of displaced workers. Similarly Fackler and Schnabel (2013) use our spin-off definitions to study the survival dynamics of spin-offs and start-ups in Germany. We expect researchers to come up with many more interesting research projects where this approach will be fruitful.

References

Abowd, J. M./*Vilhuber*, L. (2005): The sensitivity of economic statistics to coding errors in personal identifiers, Journal of Business & Economic Statistics 23 (2), 133–152.

Abowd, J. M./*Vilhuber*, L. (2011): National estimates of gross employment and job flows from the Quarterly Workforce Indicators with demographic and industry detail Journal of econometrics 161 (1), 82–99.

Baldwin, J./*Beckstead*, D./*Girard*, A. (2002): The importance of entry to Canadian manufacturing with an appendix on measurement issues, Statistics Canada Analytical Studies-Micro-Economic Analysis Division Working Paper (189).

Bartelsman, E. J./*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.

Bauer, T. K./*Schmucker*, A./*Vorell*, M. (2008): KMU und Arbeitsplatzdynamik: Eine Analyse auf Basis der Beschäftigten-Historik (SMEs and employment dynamics: an analysis based on the Employee History), Zeitschrift für ArbeitsmarktForschung-Journal for Labour Market Research 41 (2/3), 199–221.

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

Brixy, U. (2008): Welche Betriebe werden verlagert?: Beweggründe und Bedeutung von Betriebsverlagerungen in Deutschland, IAB Discussion Paper.

Brown, C./*Haltiwanger*, J./*Lane*, J. (2008): Economic turbulence: Is a volatile economy good for America?, University of Chicago Press.

Contini, B./*Revelli*, R. (2007): The process of job creation and job destruction in the Italian economy, Labour 1 (3), 121–144.

Davis, S. J./*Haltiwanger*, J. C./*Schuh*, S. (1996): Job Creation and Destruction. MIT Press.

Dundler, A./*Stamm*, M./*Adler*, S. (2007): The establishment history panel BHP 1.0: handbook version 1.0. 0, FDZ Datenreport, Documentation on Labour Market Data.

Dunne, T./*Roberts*, M. J./*Samuelson*, L. (1989): Plant turnover and gross employment flows in the US manufacturing sector, Journal of Labor Economics 7 (1), 48–71.

Eurostat/OECD (2007): Eurostat – OECD Manual on Business Demography Statistics. Luxembourg.

Fackler, D./*Schnabel*, C. (2013): Survival of spinoffs and other startups: First evidence for the private sector in Germany, 1976–2008, Institut für Wirtschaftspolitik und Quantitative Wirtschaftsforschung – Discussion Papers. 6.

Fackler, D./*Schnabel*, C./*Wagner*, J. (2012): Establishment exits in Germany: The role of size and age, Small Business Economics, 1–18.

Fort, T./*Haltiwanger*, J./*Jarmin*, R./*Miranda*, J. (2012): How Firms Respond to Business Cycles: The Role of Firm Age and Firm Size, mimeo.

Foster, L./*Grim*, C./*Haltiwanger*, J. (2013): Reallocation in the Great Recession: Cleansing or Not?, Working Paper.

Fritsch, M./*Brixy*, U. (2004): The establishment file of the German social insurance statistics, Schmollers Jahrbuch 124 (2), 183–190.

Fritsch, M./*Grotz*, R./*Brixy*, U./*Niese*, M./*Otto*, A. (2002): Gründungen in Deutschland: Datenquellen, Niveau und räumlich-sektorale Struktur. Springer.

Geurts, K./*Ramioul*, M./*Vets*, P. (2009): Using employee flows to improve measures of job creation and destruction and firm dynamics: The case of Belgium, MPRA Paper. (15306).

Haltiwanger, J./*Jarmin*, R./*Miranda*, J. (2008): Business Formation and Dynamics by Business Age: Results from the New Business Demography Statistics, Technical Report, working paper 2008.

Jacobson, L. S./*LaLonde*, R. J./*Sullivan*, D. G. (1993): Earnings Losses of Displaced Workers, American Economic Review 83 (4), 685–709.

Persson, H. (1999): Job flows and worker flows in Sweden 1986–1995, Essays on Labour Demand and Career Mobility, Swedish Institute for Social Research, Dissertation Series. 40.

Revelli, R. (1996): Statistics on job creation: issues in the use of administrative data, Job creation and loss. Analysis, policy and data development.

Schmieder, J. F. (2013): What Causes Wage Dispersion? Evidence from New Firms, mimeo, Boston University.

Schmieder, J. F./*von Wachter*, T./*Bender*, S. (2010): The long-term impact of job displacement in Germany during the 1982 recession on earnings, income, and employment, IAB Discussion Paper.

Vartiainen, J. (2004): Measuring interfirm mobility with an administrative dataset, National Institute for Economic Research. Mimeograph.

Vilhuber, L. (2009): Adjusting imperfect data: overview and case studies. Chicago.

von Wachter, T./*Song*, J./*Manchester*, J. (2009): Long-Term Earnings Losses due to Job Separation During the 1982 Recession: An Analysis Using Longitudinal Administrative Data from 1974 to 2004, mimeo.

Appendix

Table A-1

The Distribution of Clustered Worker Flows among Entering and Exiting Establishments (1976 – 2004)

Panel A: Entries

		Predecessor exits MCI / Predecessor Employment			Predecessor continues MCI / Predecessor Employment			No predecessor MCI=0
		<30%	30 – 80%	>80%	<30%	30 – 80%	>80%	
	MCI / Inflows							
≤3 empl.	-	124,863 2.63	187,893 3.95	199,348 4.19	1,076,374 22.64	181,330 3.81	43,249 0.91	2,137,606 44.96
>3 empl.	<30%	27,949 0.59	19,234 0.40	10,566 0.22	185,437 3.90	18,229 0.38	3,366 0.07	31,017 0.65
	30 – 80%	26,462 0.56	123,057 2.59	37,752 0.79	101,279 2.13	40,365 0.85	3,230 0.07	
	>80%	10,996 0.23	42,613 0.90	38,881 0.82	54,802 1.15	24,098 0.51	4,214 0.09	

Continued next page

Table A-1 (continued)

Panel B: Exits

		Successor is entrant			Successor is existing estab.			No successor
		Successor is entrant MCO / Successor Employment			Successor is existing estab. MCO / Successor Employment			No successor MCO=0
		<30%	30−80%	>80%	<30%	30−80%	>80%	
	MCO / Outflows							
≤3 empl.	-	124,863	187,893	199,348	1,076,374	181,330	43,249	2,137,606
		2.63	3.95	4.19	22.64	3.81	0.91	44.96
>3 empl.	<30%	27,949	19,23	10,566	185,437	18,229	3,366	31,017
		0.59	0.40	0.22	3.90	0.38	0.07	0.65
		<30%	30−80%	>80%	<30%	30−80%	>80%	
	30−80%	26,462	123,057	37,752	101,279	40,365	3,230	
		0.56	2.59	0.79	2.13	0.85	0.07	
	>80%	3,158	23,059	37,625	24,277	12,375	2,050	
		0.07	0.55	0.89	0.58	0.29	0.05	

Notes: The first row in each cell shows the number of establishments, the second row the percentage of the total (among entries and exits). MCI stands for Maximum Clustered Inflow: the size of the largest cluster of inflowing current workers. Inflows stands for all the total number of workers that arrived since the previous year at an EID, which for a new EID is the same as total current employment. MCO stands for Maximum Clustered Outflows: the size of the largest cluster of outflowing current workers. Outflows are all workers that leave the EID until the next year.

Table A-2

The Distribution of Establishment Entry Categories by Establishment Size in Year of Entry

Number of Employees	ID-Change	Spin-off pulled	Spin-off/pushed	New estab. (small)	New estab. (med & big)	New estab. (fuzzy)	Unclear	Total
Panel A: Number of Establishments								
≤3				3,950,679				3,950,679
4–9	23,920	40,751	32,035		223,767	189,552	27,479	537,504
10–19	8,246	17,609	11,955		45,394	60,659	9,816	153,679
20–49	4,413	12,290	6,706		20,749	30,092	5,059	79,309
50–99	1,283	4,501	1,913		4,257	7,308	1,567	20,829
100–249	754	2,584	817		1,341	2,887	849	9,232
250–499	168	736	142		221	494	252	2,013
500–999	7/	295	3/		48	137	124	710
1000+	2/	134	/		23	34	50	273
Total	38,881	78,900	53,609	3,950,679	295,800	291,163	45,196	4,754,228
Panel B: Number of Workers in Establishment Type								
≤3				4,990,187				4,990,187
4–9	134,527	235,190	186,434		1,191,253	1,075,007	160,023	2,982,434
10–19	108,725	235,982	157,679		601,661	809,100	131,756	2,044,903
20–49	131,382	371,269	200,541		605,412	887,678	150,062	2,346,344
50–99	87,753	310,513	129,741		282,668	493,186	107,708	1,411,569
100–249	111,644	388,131	119,620		194,604	423,171	128,156	1,365,326
250–499	57,311	252,191	48,499		74,517	164,336	87,939	684,793
500–999	49,022	198,914	20,089		30,207	89,934	85,822	473,988
1000+	30,994	230,378	/		/	54,115	88,461	471,122
Total	711,358	2,222,568	883.6//	4,990,187	3.026.4//	3,996,527	939,927	16,770,666

Note: Data confidentiality rules prohibit the publication of table cells with less than 20 observations. For this reason cells with less than 20 observations have been replaced by "/". Furthermore certain digits in the total counts have similarly been replaced by "/" to make it impossible to infer the cell counts indirectly.

Table A-3

The Distribution of Establishment Exit Categories by Establishment Size in Year prior to Exit

Number of Employees	ID-Change	Takeover/ Restructuring	Spin-off/ pushed	Small Death	Atomized Death	Fuzzy Death	Unclear	Total
Panel A: Number of Establishments								
≤3				3,494,502				3,494,502
4–9	23,094	21,589	51,890		205,728	155,387	17,128	474,816
10–19	8,118	8,044	17,449		53,866	50,846	6,480	144,803
20–49	4,211	4,741	10,954		25,770	24,875	3,244	73,795
50–99	1,247	1,416	3,791		5,645	5,700	860	18,659
100–249	701	661	1,882		1,770	2,180	393	7,587
250–499	166	153	364		276	413	108	1,480
500–999	7/	3/	99		6/	86	4/	396
1000+	/	/	22		/	32	/	105 105
Total	37,625	36,652	86,451	3,494,502	293,127	239,519	28,267	4,216,143
Panel B: Number of Workers in Establishment Type								
≤3				4,321,132				4,321,132
4–9	130,837	122,783	297,728		1,121,023	887,946	99,784	2,660,101
10–19	106,760	106,881	235,032		718,318	673,998	84,519	1,925,508
20–49	125,143	141,657	329,951		752,926	729,853	95,700	2,175,230
50–99	85,551	97,192	260,665		379,241	384,408	58,061	1,265,118
100–249	105,547	97,431	277,574		251,065	317,098	57,725	1,106,440
250–499	56,632	50,630	122,586		91,457	136,819	36,885	495,009
500–999	46,526	26,592	65,339		39,727	57,362	27,954	263,500
1000+	24,144	18,313	40,032		23,385	59,778	19,284	184,936
Total	681,140	661,479	1,628,907	4,321,132	3,377,142	3,247,262	479,912	1.44e+07

Note: Data confidentiality rules prohibit the publication of table cells with less than 20 observations. For this reason cells with less than 20 observations have been replaced by "/". Furthermore certain digits in the total counts have similarly been replaced by "/" to make it impossible to infer the cell counts indirectly.