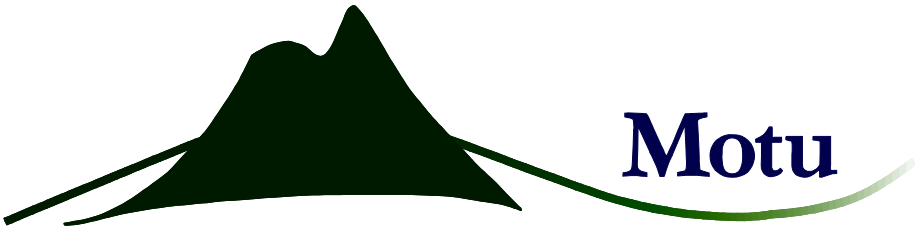


Keeping it Together: Tracking Firms in New Zealand's Longitudinal Business Database

Richard Fabling*

Motu Working Paper 11-01
Motu Economic and Public Policy Research
March 2011

* Richard Fabling is an Affiliate of Motu, and is a Senior Adviser (Research) in the Economics Department at the Reserve Bank of New Zealand

Author contact details

Richard Fabling
Reserve Bank of New Zealand
richard.fabling@rbnz.govt.nz

Acknowledgements

Thanks to Dave Maré and Lynda Sanderson for helpful comments. This analysis forms part of the author's secondment agreement with Statistics New Zealand.

Disclaimer

This research was undertaken while Richard Fabling was on secondment to Statistics New Zealand. Access to the data used in this study was provided by Statistics NZ in accordance with the security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Act are allowed to see data about a particular business or organisation. The results of this work have been confidentialised to protect individual businesses from identification. The opinions, findings, recommendations, and conclusions expressed in this report are those of the authors. Statistics New Zealand, the Reserve Bank of New Zealand, and Motu take no responsibility for any omissions or errors in the information contained here.

The results are based in part on tax data supplied by Inland Revenue to Statistics New Zealand under the Tax Administration Act 1994. This tax data must be used only for statistical purposes, and no individual information is published or disclosed in any other form, or provided back to Inland Revenue for administrative or regulatory purposes. Any person who had access to the unit-record data has certified that they have been shown, have read and have understood section 81 of the Tax Administration Act 1994, which relates to privacy and confidentiality. Any discussion of data limitations or weaknesses is not related to the data's ability to support Inland Revenue's core operational requirements.

Motu Economic and Public Policy Research

PO Box 24390
Wellington
New Zealand

Email info@motu.org.nz
Telephone +64 4 9394250
Website www.motu.org.nz

 ISSN 1176-2667 (Print), ISSN 1177-9047 (Online).

Abstract

Having good longitudinal identifiers is important in empirical microeconomics, since researchers often need to be able to observe the same unit over time to make causal inferences. However, firm identifiers in Statistics New Zealand's Longitudinal Business Database can be "broken" by, among other things, changes in the legal status of the firm. This paper proposes a simple method for repairing broken firm identifiers, making use of existing plant migration data. We show that making such repairs materially reduces the apparent rate of business entry and exit, and allows real economic phenomena, such as small business incorporation, to be observed for the first time.

JEL codes

D20 - Production and Organizations: General

Keywords

Firm identifiers; entry and exit; plant transfers

Summary Haiku

gone are the workers
together toiling elsewhere
thus the firm lives on

Contents

1. Motivation

Economists interested in analysing firms in a panel setting (i.e., over time) should be concerned with the proper identification of entering, ongoing, and ceasing businesses. Causal inference in empirical microeconomics often relies on observing the same unit (firm, plant, worker, household, et cetera) before, during, and after some event.

In New Zealand the pre-eminent business research database is the prototype Longitudinal Business Database (LBD) maintained by Statistics New Zealand (Fabling, 2009). The LBD has two primary units of observation, the plant (geographic unit) and the firm (enterprise), with the latter being the filing unit for most data.

In the LBD, longitudinal characteristics of firms and plants (e.g., industry) are derived by unwinding changes made to Statistics New Zealand's Business Frame (BF), which provides accurate point-in-time representations of the population of firms. BF identifiers track legal entities over time, which is entirely satisfactory for the purpose of compiling point-in-time statistics or for drawing samples for cross-sectional business surveys. However continuity of legal entities and firms is not always the same thing. For example, a sole proprietor may decide to incorporate their business, while continuing to employ the same staff in the same location, producing the same goods and services. This business may be represented in the LBD as two firms – one exiting, one entering – where an economist would say there is one ongoing firm.

This paper provides a method for repairing "broken" firm identifiers (IDs), by making use of the effort Statistics New Zealand has put into maintaining true longitudinal plant-level IDs. The next section outlines the methodology and quantifies the effect of repairs with some simple statistics, while section three briefly summarises. The code to repair firm IDs is provided in the Appendix.

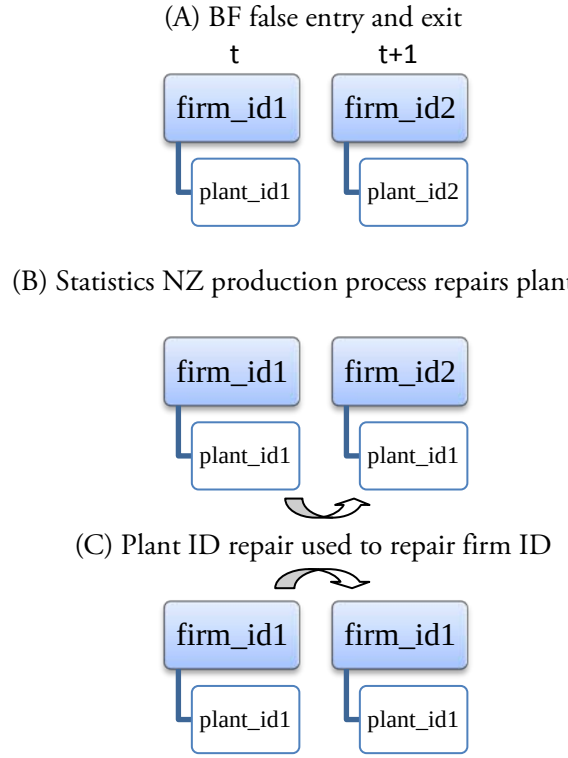
2. Repairing firm identifiers

2.1. Methodology

Consider the simplest instance of a broken firm ID, where a single location (plant) firm continues to operate, but for some reason is treated as exiting (Figure 1, panel A). Evidence of the broken firm ID comes from Statistics New Zealand processes which identify continuing plants and repair the plant ID

(panel B).[1] The majority of plant ID repairs are identified through BF maintenance procedures (i.e., manual operator investigation), while a smaller number are picked up through tax ID, location, activity, and employee-matching methods. This paper outlines a simple methodology for using repaired plant IDs to identify and fix broken firm IDs (panel C).[2]

Figure 1 – A broken firm ID identified and repaired



The firm ID repair process is kept simple by only addressing repairs that fit the mould of Figure 1 (allowing for the possibility that a firm may have multiple employing plants). Specifically, broken firm IDs that are repaired satisfy the following criteria:

1. firm_id1 (the source ID) employs up to month *t* and not after;
2. firm_id2 (the target ID) employs from month *t* +1 and not before;

[1] We use the terms firm ID and plant ID for simplicity. Statistics New Zealand call firm IDs "enterprise numbers" (ENTs) and plant IDs "geographic numbers" before, and "permanent business numbers" (PBNs) after, repairs.

[2] Undoubtedly there are instances where firm IDs have been broken that are not identifiable through plant ID repairs. In particular, firms with very low employment are excluded from matching procedures, and are not the primary focus of manual BF maintenance. Data held outside the LBD would presumably be required to aid repair to those firm IDs. Such a data-gathering exercise lies outside the scope of this paper.

3. the source ID doesn't have another target id in month $t+1$; and
4. the target ID doesn't have another source id in month t.

The first two criteria ensure that there is no overlapping employment period between firm_id1 and firm_id2 – an event that couldn't happen if the IDs represented the same firm. These rules exclude partial sales of firms (e.g., sales of "going concern" plants). Criteria three and four exclude other complex transactions (e.g., group mergers or buyouts) that would require additional assumptions or manual examination to establish whether a continuing firm exists.

Table 1 – Repair candidates by outcome of repair process

	Source IDs	Proportion
Not resulting in a firm ID repair	17,922	0.391
Multiple months	1,221	0.027
Multiple sources/targets within month	1,530	0.033
Source and target both employing before and after	3,075	0.067
Source continues employing after	6,750	0.147
Target was already employing before	5,346	0.117
Resulting in a firm ID repair	27,942	0.609
Total distinct source IDs with at least one target ID	**45,864**	

Table 1 provides a count of repair candidates in the LBD (distinct firm IDs that pass an employing plant to another firm ID), together with the proportion of repaired IDs and a classification of those not repaired.[3] Overall, three fifths of candidates are unambiguously incorrect breaks in firm IDs and are repaired under the simple rules. The classification of the remaining 39% of candidates is hierarchical to avoid double-counting.[4] Violations of the employment continuity rules (rules one and two) are the primary reason why plant ID repairs are not used to fix firm IDs. At least a third of candidates are not

[3] These statistics are drawn from the latest (August 2010) archive version of the LBD, which includes eight longitudinal business frame years (2000–2007). The code appendix is written to work on any version of the LBD. All counts are random-rounded (base three) in compliance with Statistics NZ confidentiality rules.

[4] For example, most multiple month transfers involve multiple target IDs. As a consequence, relaxing the simple rule to allow transitions to occur over two months (or any other period) would not materially increase the number of firm ID repairs, but would increase the complexity of the code.

repaired (85% of non-repairs) because employment activity is maintained in the source and/or target ID where this should not occur if the source and target ID both represented the same firm.

Fabling and Grimes (2009) take a different approach to repairing broken firm IDs. Since their analysis depends on a smaller panel of firms,[5] they take all pairs of firms linked by plant transfers between 2001 and 2005, and then manually inspect names, addresses, detailed industry, and employment to find and repair broken firm IDs. Their match rate from this labour-intensive approach is 58%, remarkably similar to the rules-based approach outlined here.

Thus it seems likely that the method in this paper repairs the bulk of IDs that are identifiable as repair candidates and which actually represent undesirable breaks in firm id continuity. Following Statistics NZ's classification of firm IDs as ENTs, we refer to the repaired firm IDs as PENTs (Permanent ENTerprises).

2.2. Are repaired IDs better?

How might we verify that the use of PENTs represents an improvement over the status quo? The earlier example of incorporation sheds light on this question. After incorporation, a firm is obliged to start filing company income tax returns (IR4s).[6] In the absence of PENTs this known economic phenomenon is almost never observed – using unrepaired firm IDs, a mere 0.005% of employing sole proprietors and partnerships transition into IR4 filing (a total of 18 firms over seven years). By comparison, using PENTs the transition rate is 2.4%, or around 9,000 firms over seven years. So, the use of PENTs prevents a sizeable and biased (owners have chosen to change legal form) proportion of employing businesses from appearing to exit and enter the population.

Across all firms, entry and exit rates in and out of employing are one and a half percentage points lower using PENTs instead of unrepaired firm IDs. Entry (exit) rates are 14.3% (12.2%) using PENTs, and 15.8% (13.7%) using unrepaired firm IDs. Balanced panels for key longitudinal datasets are also larger using PENTs. For example, the 2000–2008 Annual Enterprise Survey balanced panel is almost 2% larger.

[5] Being respondents to both the 2001 Business Practices Survey and the 2005 Business Operations Survey. By sample design, this panel includes only private-for-profit firms with six or more employees.

[6] If incorporation leads to a break in a firm ID, then looking at changes in the recorded business type of the firm will not tell us anything useful. However, assuming (in cross-section) that the business type indicator on the BF is accurate, then the IR4 "response rate" among employing limited liability companies is around 92% (and 0.1% for sole proprietors and partnerships). Actual compliance is likely higher, implying that IR4 filing is an accurate indicator of limited liability company status.

Overall, repaired ID firms constitute 7.2% of ever-employing firms and, since firms with repaired IDs employ, on average, for more years (seven versus four years for other firms), 11.9% of firm-year observations. Consistent with small firm incorporation being a major reason for breaks in firm IDs, repaired firms are on average smaller in employment terms, having an average 7.9 rolling-mean employment compared to 10.6 for other firms.

2.3. Filing patterns, or, "You're still moving? I thought you were dead!"

Another lens on the validity of repairing firm IDs comes from examining filing of Goods and Services Tax (GST) returns and IR10 accounts. In employing years, firms with repaired IDs have comparable tax filing rates to other firms for reported income or expenditure, or total fixed assets.[7] If source firm IDs consistently continued to file, say, GST returns, after they had passed their employing plants to the target firm ID we might question whether the source and target IDs really belonged to the same firm. Mitigating this interpretation is the fact that many firms continue to file tax returns long after they cease employing. For this reason, we benchmark the filing patterns of repaired ID firms against all other firms.

Figures two and three present results for income/expenditure and total fixed assets respectively. The vertical axis is the proportion of firms filing the relevant return, while the horizontal axis is the number of years before (after) the firm/ID starts (stops) employing (t = 0 pools years where the firm is employing). For repaired IDs, therefore, observations with negative values of t are target IDs before they employ, while positive values of t are associated with source IDs that have ceased employing. For all other firms negative (positive) t is associated with years prior to the first (after the last) year that the firm ever employs.[8] To abstract from other determinants of response rates, the analysis focuses on firms that ever employ and that ever file the relevant tax form.

Figures two and three show that targets in repaired ID firms are seldom associated with filed tax returns before the year they start employing (negative t). This is not the case for other firm IDs and likely reflects the fact that most breaks in firm IDs are associated with the creation of new legal entities and, therefore,

[7] Income/expenditure filing is defined as positive GST sales or purchases from the Business Activity Indicator data, or positive total income or total expenditure from an edit-checked front page of an IR10 return. Total fixed assets come from edit-checked IR10 back pages.

[8] These results are robust to estimating over a shorter ($-7<t<7$) time frame, suggesting that mismeasurement of the first/last employment year due to censoring is not an issue. The estimate of t = 0 filing rates for repaired IDs is biased downwards by the choice to count both the source and target responses independently in years of plant handover. The alternative of counting filings jointly (i.e., counting once in the denominator, and once in the numerator if either ID files) produces a higher filing rates at t = 0, but still similar to other firms.

new Inland Revenue Department (IRD) numbers linked to – and initiating the creation of – the target firm ID. Source ID filing after ceasing employment (positive *t*) is a more common phenomenon for both income/expenditure and fixed assets, though still substantially less common than for other firms.

Figure 2 – Income/expenditure filing patterns

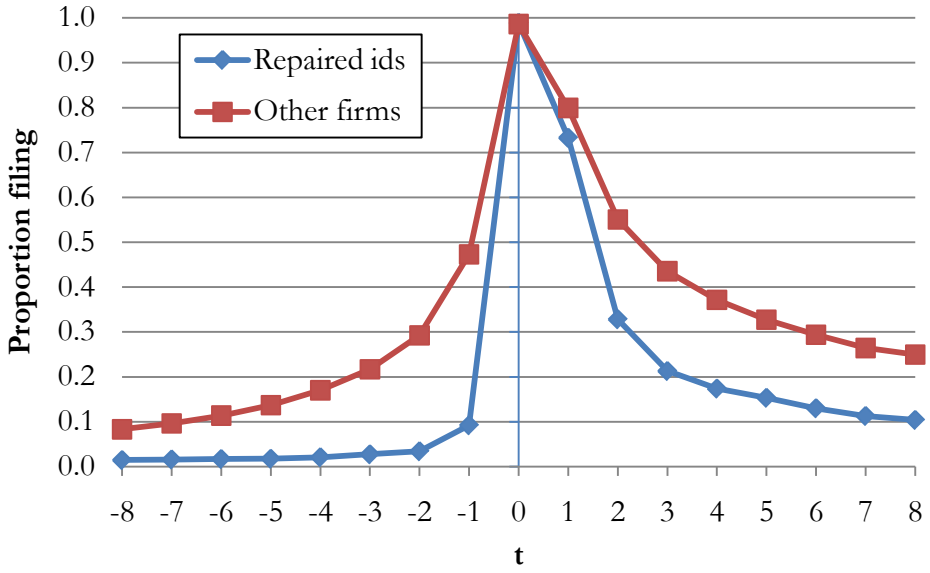
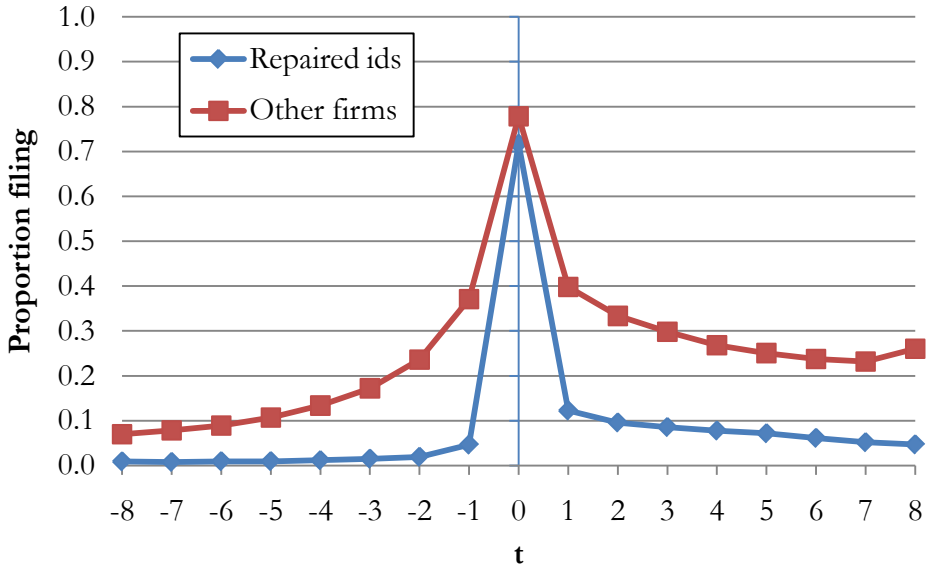
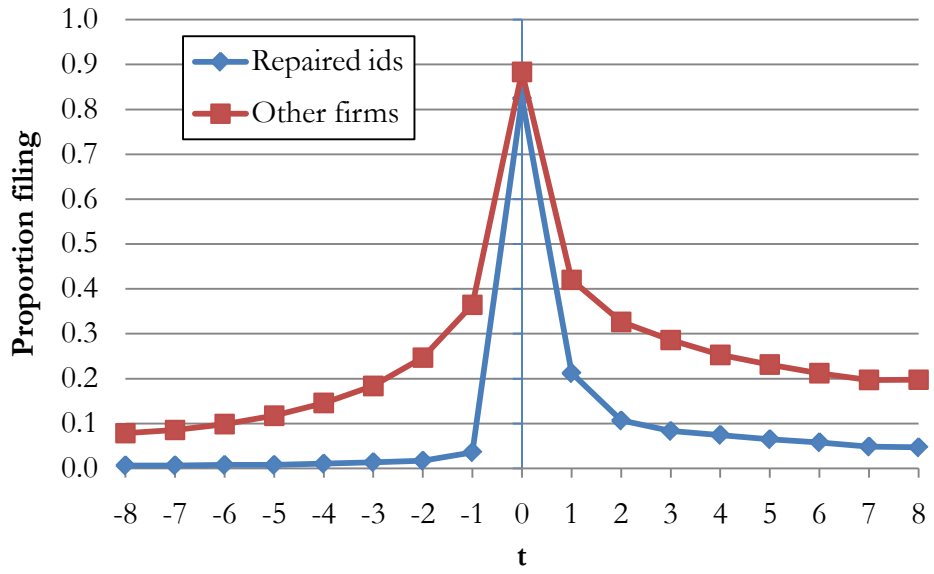


Figure 3 – Total fixed assets filing patterns



One reason filing may persist is because firms have not truly ceased operating, but instead have ongoing working proprietor labour input not included in the employee count.[9] Figure four shows working proprietor filing using the same methodology as for other tax filing. Conditional on ever filing a tax return indicative of working proprietors, many firms have working proprietors before and after they cease having employees. As with other filing patterns, repaired ID firms have almost no related filing for targets before plant transfer and substantially lower rates of working proprietor reporting in source IDs after transfer.

Figure 4 – Working proprietor filing patterns



Furthermore, these tax returns do not necessarily imply labour input – they could simply reflect residual payments to owners. In other words, they could be another example of the issue of late filing, rather than the root explanation. In research using the LBD, working proprietor filing is usually assumed to indicate labour input and this is probably generally the case since a sizeable share of the workforce is self-employed and many firms would otherwise appear to generate output in the absence of any labour input. However, this assumption may be less applicable in the scenario we examine here where firms transition out of having employing plants (repaired IDs), or employees (all other firms). In any event, at best working proprietor filing rates "explain" around 40% (50%) of post-employment income/expenditure reporting by source IDs (other firms).

[9] Because the owner is not paid a pay-as-you-earn (PAYE) wage.

In summary, the pre-employment filing patterns for target firm IDs are highly consistent with the repairs implemented, while the filing patterns post-employment for source firm IDs are less so. Compared to the filing patterns of other firms, sources are more likely to appear to be inactive, though users of PENTs will have to make important decisions whether to treat post-employment filing by source IDs as part of an aggregated PENT-level return (for example, Fabling and Maré (forthcoming) choose not to include IR10 returns in non-employing years). To aid users in implementing their choice, the code in the Appendix segments time into (PAYE) employment periods for repaired ID firms, so that researchers can easily identify pre- and post-employment periods for source and target IDs.

3. Conclusions

This paper proposes a simple method for repairing broken firm IDs, making use of existing plant migration data. The method sets aside more complex plant transfers that likely require additional (non-LBD) data to allow identification of repairable firm IDs. Despite this, the majority of candidate breaks are fixed by the proposed method, decreasing the estimated entry and exit rates of firms by one and a half percentage points. Furthermore, using repaired IDs produces results consistent with economic intuition. In particular, small firms are observed to incorporate, where previously they were not. Target firm IDs are also highly unlikely to be associated with filed tax returns prior to being associated with employing plants, consistent with their being a new legal form (with new IRD numbers) of an ongoing firm. Post-employment filing associated with source firm IDs is still common, and theoretically inconsistent with the business moving to a new legal form with a new IRD number, though instances of this filing decay far more quickly than for the general population of firms. Users of PENTs will need to decide for themselves whether to treat this filing as part of a consolidated return for the firm.

4. References

Fabling, Richard. 2009. "A Rough Guide to New Zealand's Longitudinal Business Database," *Global COE Hi-Stat Discussion Paper 103*, Tokyo: Hitotsubashi University.

Fabling, Richard and Arthur Grimes. 2009. "The 'Suite' Smell of Success: Complementary Personnel Practices and Firm Performance," *RBNZ Discussion Paper 2009/13*, Wellington: Reserve Bank of New Zealand.

Fabling, Richard and David C. Maré. Forthcoming. "Production Function Estimation Using New Zealand's Longitudinal Business Database," *Motu Working Paper*.

5. Code appendix

```
-----------------------------------------------------------------
--Code to repair enterprise id breaks using PBN migration
--Author: Richard Fabling
--Database: ibuldd_clean(_archive_*)
-----------------------------------------------------------------

--Collect all employing pbns that change enterprises
create table #pbn_links
        (pbn_nbr                char(10)        NOT NULL,
         source_month   int             NOT NULL,
         target_month   int             NOT NULL,
         source_ent     char(10)        NOT NULL,
         target_ent     char(10)        NOT NULL,
         primary key clustered (pbn_nbr, source_month, target_month))

insert into #pbn_links(pbn_nbr, source_month, target_month,
                                source_ent, target_ent)
select
p1.pbn_nbr,
source_month=p1.dim_month_key,
target_month=p2.dim_month_key,
source_ent=p1.enterprise_nbr,
target_ent=p2.enterprise_nbr
from load_lbf_fact_pbn_employee_count p1
join load_lbf_fact_pbn_employee_count p2
on p1.pbn_nbr=p2.pbn_nbr
and p1.enterprise_nbr<>p2.enterprise_nbr
and dbo.fn_month_get_next(p1.dim_month_key,1)=p2.dim_month_key
where p1.leed_employee_count_nbr>0
and p2.leed_employee_count_nbr>0

--Find enterprises that satisfy repair rules
create table #repairs
        (source_month   int             NOT NULL,
         source_ent     char(10)        NOT NULL,
         target_ent     char(10)        NOT NULL,
         primary key clustered (source_month, source_ent))

insert into #repairs(source_month, source_ent, target_ent)
select distinct
source_month,
pbn.source_ent,
pbn.target_ent
from #pbn_links pbn
--Source never has PBN employment after transfer
where source_ent not in (select
                                source_ent
```

```
                                        from #pbn_links lnk
                                        join load_lbf_fact_pbn_employee_count pbn
                                        on source_ent=pbn.enterprise_nbr
                                        and lnk.target_month<=pbn.dim_month_key
                                        where leed_employee_count_nbr>0)

--Target never has PBN employment before transfer
and target_ent not in (select
                                        target_ent
                                        from #pbn_links lnk
                                        join load_lbf_fact_pbn_employee_count pbn
                                        on target_ent=pbn.enterprise_nbr
                                        and lnk.source_month>=pbn.dim_month_key
                                        where leed_employee_count_nbr>0)
--Source doesn't have another target in that month
and not exists (select
                        1
                        from #pbn_links ps
                        where ps.source_ent=pbn.source_ent
                        and ps.source_month=pbn.source_month
                        and ps.target_ent<>pbn.target_ent)
--Target doesn't have another source in that month
and not exists (select
                        1
                        from #pbn_links pt
                        where pt.target_ent=pbn.target_ent
                        and pt.target_month=pbn.target_month
                        and pt.source_ent<>pbn.source_ent)

--Create map between PENTs and enterprise_nbrs
create table #pent
        (pent                   char(10)        NOT NULL,
         enterprise_nbr  char(10)        NOT NULL,
         start_month     int             NOT NULL,
         end_month               int             NOT NULL,
         primary key clustered (pent, enterprise_nbr))

DECLARE @first_emp_mth int
select @first_emp_mth=MIN(dim_month_key)
from load_lbf_fact_pbn_employee_count
where leed_employee_count_nbr>0

DECLARE @last_emp_mth int
select @last_emp_mth=MAX(dim_month_key)
from load_lbf_fact_pbn_employee_count
where leed_employee_count_nbr>0

--Start with first enterprise number in chain
--ie, where source_ent is not a target_ent
```

```
--PENT will be the first enterprise_nbr in sequence
insert into #pent(pent, enterprise_nbr, start_month, end_month)
select
pent=source_ent,
enterprise_nbr=source_ent,
start_month=@first_emp_mth,
end_month=source_month
from #repairs
where source_ent not in (select target_ent from #repairs)



--Now loop over subsequent enterprise_nbrs in the chain
DECLARE @mth int
select @mth=MIN(source_month) from #repairs

DECLARE @last_mth int
select @last_mth=MAX(source_month) from #repairs

while @mth<=@last_mth
begin
--Truncate link period end for source enterprise
 update #pent
 set end_month=@mth
 where enterprise_nbr in (select source_ent
                                                from #repairs
                                                where source_month=@mth)

--Set target PENT to source PENT
 insert into #pent(pent, enterprise_nbr, start_month, end_month)
 select
 pent=pen.pent,
 enterprise_nbr=target_ent,
 start_month=dbo.fn_month_get_next(@mth,1),
 end_month=@last_emp_mth
 from #repairs rep
 join #pent pen
 on rep.source_ent=pen.enterprise_nbr
 where source_month=@mth

 set @mth=dbo.fn_month_get_next(@mth,1)
end

--Complete table by adding single-enterprise_nbr PENTs
insert into #pent(pent, enterprise_nbr, start_month, end_month)
select
pent=enterprise_nbr,
enterprise_nbr,
```

```
start_month=@first_emp_mth,
end_month=@last_emp_mth
from fact_lbf_enterprise_year
where enterprise_nbr not in (select enterprise_nbr from #pent)
group by enterprise_nbr
```

Recent Motu Working Papers

All papers in the Motu Working Paper Series are available on our website www.motu.org.nz, or by contacting us on info@motu.org.nz or +64 4 939 4250.

11-02 Roskruge, Matthew, Arthur Grimes, Philip McCann and Jacques Poot. “Homeownership and Social Capital in New Zealand”.

11-01 Fabling, Richard. 2011. “Keeping It Together: Tracking Firms on New Zealand’s Longitudinal Business Database”.

10-14 Stroombergen, Adolf. 2010 “The International Effects of Climate Change on Agricultural Commodity Prices, and the Wider Effects on New Zealand”.

10-13 Olssen, Alex; Hugh McDonald, Arthur Grimes, and Steven Stillman. 2010. “A State Housing Database: 1993–2009”.

10-12 Cochrane, William; Arthur Grimes; Phillip McCann and Jacques Poot. 2010. “The Spatial Impact of Local Infrastructural Investment in New Zealand”.

10-11 Grimes, Arthur and Chris Young. 2010. “Anticipatory Effects of Rail Upgrades: Auckland’s Western Line”.

10-10 Grimes, Arthur and Chris Young. 2010. “A Simple Repeat Sales House Price Index: Comparative Properties Under Alternative Data Generation Processes”.

10-09 Evans, Lewis; Greame Guthrie and Andrea Lu. 2010. "A New Zealand Electricity Market Model: Assessment of the Effect of Climate Change on Electricity Production and Consumption".

10-08 Coleman, Andrew and Hugh McDonald. 2010. "“No country for old men”: a note on the trans-Tasman income divide".

10-07 Grimes, Arthur; Mark Holmes and Nicholas Tarrant. 2010. "New Zealand Housing Markets: Just a Bit-Player in the A-League?"

10-06 Le, Trinh; John Gibson and Steven Stillman. 2010. "Household Wealth and Saving in New Zealand: Evidence from the Longitudinal Survey of Family, Income and Employment".

10-05 Grimes, Arthur. 2010. "The Economics of Infrastructure Investment: Beyond Simple Cost Benefit Analysis".

10-04 van Benthem, Arthur and Suzi Kerr. 2010. “Optimizing Voluntary Deforestation Policy in the Face of Adverse Selection and Costly Transfers”.

10-03 Roskruge, Matthew; Arthur Grimes; Philip McCann and Jacques Poot. 2010. "Social Capital and Regional Social Infrastructure Investment: Evidence from New Zealand”.

10-02 Coleman, Andrew. 2010. "Uncovering uncovered interest parity during the classical gold standard era, 1888-1905".

10-01 Coleman, Andrew. 2010. "Squeezed in and squeezed out: the effects of population ageing on the demand for housing".

09-17 Todd, Maribeth and Suzi Kerr. 2009. "How Does Changing Land Cover and Land Use in New Zealand relate to Land Use Capability and Slope?"

09-16 Kerr, Suzi and Wei Zhang. 2009. "Allocation of New Zealand Units within Agriculture in the New Zealand Emissions Trading System”.

09-15 Grimes, Arthur; Cleo Ren and Philip Stevens. 2009. "The Need for Speed: Impacts of Internet Connectivity on Firm Productivity”.

09-14 Coleman, Andrew and Arthur Grimes. 2009. "Fiscal, Distributional and Efficiency Impacts of Land and Property Taxes”.

09-13 Coleman, Andrew. 2009. "The Long Term Effects of Capital Gains Taxes in New Zealand”.