Employment misclassification in survey and administrative reports

Dean Hyslop and Wilbur Townsend

Motu Working Paper 16-19

Motu Economic and Public Policy Research

October 2016

Document information

Author contact details

Dean Hyslop

Motu Economic and Public Policy Research

dean.hyslop@motu.org.nz

Wilbur Townsend

Motu Economic and Public Policy Research

wilbur.townsend@motu.org.nz

Acknowledgements

This research was supported in part by the Royal Society of New Zealand Marsden Fund Grant MEP1301. We are grateful to Corey Allan for his assistance constructing the data extract for the analysis.

Disclaimer

The results in this working paper are not official statistics, they have been created for research purposes from the Integrated Data Infrastructure (IDI), managed by Statistics New Zealand. The opinions, findings, recommendations, and conclusions expressed in this working paper are those of the authors, not Statistics NZ, or Motu Economic and Public Policy Research. Access to the anonymised data used in this study was provided by Statistics NZ in accordance with security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business, or organisation, and the results in this working paper have been confidentialised to protect these groups from identification. Careful consideration has been given to the privacy, security, and confidentiality issues associated with using administrative and survey data in the IDI. Further detail can be found in the Privacy impact assessment for the Integrated Data Infrastructure available from www.stats.govt.nz. The results are based in part on tax data supplied by Inland Revenue to Statistics NZ under the Tax Administration Act 1994. This tax data must be used only for statistical purposes, and no individual information may be published or disclosed in any other form, or provided to Inland Revenue for administrative or regulatory purposes. Any person who has 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 secrecy. Any discussion of data limitations or weaknesses is in the context of using the IDI for statistical purposes, and is not related to the data's ability to support Inland Revenue's core operational requirements.

Abstract

This paper analyses measurement error in the classification of employment. We show that the true employment rate and time-invariant error rates can be identified, given access to two measures of employment with independent errors. Empirical identification requires at least two periods of data over which the employment rate varies. We estimate our model using matched survey and administrative data from Statistics New Zealand's Integrated Data Infrastructure. We find that both measures have error, with the administrative data being substantially more accurate than the survey data. In both sources, false positives are much more likely than false negatives. Allowing for errors in both sources substantially affects estimated employment rates.

JEL codes

C18, J6, J21

Keywords

Unemployment rate, measurement error, validation study

Summary haiku

Surveys have errors.
Tax data errors are rare,
but they still matter.

Motu Economic and Public Policy Research

PO Box 24390 info@motu.org.nz +64 4 9394250
Wellington www.motu.org.nz
New Zealand

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

Contents

List of Figures

List of Tables

I. Introduction

Measurement error impedes analysis of survey data (Bound, Brown, & Mathiowetz, 2001), and a substantial literature has found measurement error in employed workers' earnings.[1] However there is little research on the mismeasurement of employment itself. This paper studies the misclassification error using two measures of employment from matched survey and administrative data.

The lack of existing research is surprising. Employment dynamics are regularly studied, and these studies require assumptions about error in their data.[2] Moreover, studies of earnings error which omit the unemployed may be biased for the same reason that estimates which omit censored observations are generally biased: they select on a non-random subsample.[3] Finally, the restrictions required to identify measurement error in employment are relatively weak, and the nature of employment measurement error has obvious implications for that in earnings.

The existing literature on error in survey data employment has lacked a distinct second measure of employment. Poterba and Summers (1986) analyse reported employment status in the US Current Population Survey using a validation reinterview, assuming that correct employment status is established during the reinterview. They estimate that 5% of non-employed individuals report being employed (are false positives), and that 2% of those employed report being not employed (are false negatives).

Alternatively, Keane and Sauer (2009) estimate a dynamic model of female employment in which employment is measured with error. By restricting the longitudinal distribution of employment they identify error rates. They estimate false positive rates of 6% - 8%, and false negative rates of about 1%. Similarly, Feng and Hu (2013) longitudinally restrict the Current Population Survey, requiring for example that next-period employment is independent of lagged employment conditional on current employment

[1]See for example Bound and Krueger (1991); Pischke (1995); Kapteyn and Ypma (2007); Abowd and Stinson (2013); Hyslop and Townsend (2016).

[2]Most studies assume their data lack error, though some have calibrated models to presumed error rates (Poterba & Summers, 1995).

[3]For example, Pischke (1995) found transitory income shocks were under-reported in surveys. The negative correlation between measurement error and transitory shocks could be produced by data in which transitory shocks are not under-reported, if observations with both negative transitory shocks *and* negative measurement error are censored and thus omitted.

and demographics. They find a false positive rate of roughly 5% and a false negative rate of roughly 2%.[4]

As far as we know, this is the first study of misclassification errors using discrepancies between two measures of employment.[5] Our data comes from Statistics New Zealand's Integrated Data Infrastructure (IDI), using a matched sample from the Survey of Family, Income and Employment (SoFIE) linked to administrative earnings records. The first employment measure is derived from SoFIE, while the second measure is derived from the administrative data. While both SoFIE and the administrative data may contain error, the causes of error in each are distinct and thus are likely independent.

We show that the true employment rate and each measure's misreporting rates can be identified, provided that the errors are independent across the two measures and that the error rates are constant over time. We find both SoFIE and the administrative data contain error. The SoFIE false positive rate is between 10% and 16%, while the false negative rate is about 3%. There is less error in the administrative data – less than 3% false positives, and between 1% and 4% false negatives – but allowing for erroneous administrative data does substantially increase estimated employment, by about 2 percentage points.

In Section II we present our model and discuss identification. In Section III we present our data and its summary statistics. In Section IV we estimate our model and present its results. Section V concludes.

II. A model of employment misclassification error

In this section we model two measures of employment. Our model is identified by three sets of assumptions: false positive and false negative rates are constant over time, are independent across the two measures, and the probability of employment is increasing in both measures. Empirical identification also requires at least two time

[4]Like Poterba and Summers (1986), Feng and Hu (2013) distinguish between the unemployed and those not in the labour force. Our data cannot make that distinction. These papers also study employment at a point in time, whereas we consider a person employed if they work at all within a year.

[5]While existing papers have omitted employment in their more formal estimates, some have included employment discrepancy summary statistics (Kapteyn & Ypma, 2007).

periods and a time-varying true employment rate.

Let E_t be the binary event that a person is employed in period t, E_t^c be the complementary event that the person is not employed, and let S_t and A_t be the events that the person is reported as employed in survey and administrative data respectively. The probabilities $P(S_t)$, $P(A_t)$ and $P(S_t, A_t)$ can be estimated with sample proportions. Our aim is to estimate the employment rate ($P(E_t)$), and the false positive $P(S_t|E_t^c), P(A_t|E_t^c)$ and false negative $(1 - P(S_t|E_t)), (1 - P(A_t|E_t))$ rates associated with each measure.

Local identification requires two sets of restrictions. First assume that the false positive and false negative rates are constant over time. For all t:

$$\begin{aligned} P(S_t|E_t) &= P(S|E), \\ P(S_t|E_t^c) &= P(S|E^c), \\ P(A_t|E_t) &= P(A|E), \\ P(A_t|E_t^c) &= P(A|E^c). \end{aligned} \tag{1}$$

This assumption implies that changes in the employment rate are the only source of year-to-year changes in reporting.

Second, assume that the false positive and false negative rates are independent across the two measures of employment.[6] Given equation (1),

$$\begin{aligned} P(S, A|E) &= P(S|E) \cdot P(A|E), \\ P(S, A|E^c) &= P(S|E^c) \cdot P(A|E^c). \end{aligned} \tag{2}$$

Given these two assumptions, with T periods there are $3T$ sample moments $\{P(S_t), P(A_t), P(S_t, A_t); t = 1, ..., T\}$ and $(4 + T)$ parameters $\{P(S|E), P(S|E^c), P(A|E), P(A|E^c), P(E_t); t = 1, ..., T\}$. By applying the law of total probability to each sample

[6] If we allow arbitrary correlation between the errors – or, equivalently, attempt to estimate the joint error terms $P(S, A|E), P(S, A|E^c)$ – the model is unidentified. With $T = 3$ periods the model has as many parameters as moment conditions, but the Jacobian of the moment conditions has determinant 0. The assumption of time-constant error rates can be relaxed when $T > 2$, which we explore in the empirical analysis below.

moment, we express them in terms of the parameters:

$$\begin{aligned} P(S_t) &= P(S|E) \cdot P(E_t) + P(S|E^c) \cdot [1 - P(E_t)], \\ P(A_t) &= P(A|E) \cdot P(E_t) + P(A|E^c) \cdot [1 - P(E_t)], \\ P(S_t, A_t) &= P(S, A|E) \cdot P(E_t) + P(S, A|E^c) \cdot [1 - P(E_t)] \\ &= P(S|E) \cdot P(A|E) \cdot P(E_t) + P(S|E^c) \cdot P(A|E^c) \cdot [1 - P(E_t)]. \end{aligned} \tag{3}$$

As specified, the model is locally just-identified when $T = 2$. The Jacobian of the moment conditions is square, with determinant

$$\Delta = [P(S|E) - P(S|E^c)]^2 \cdot [P(A|E) - P(A|E^c)]^2 \cdot [P(E_1) - P(E_2)]^2 \tag{4}$$

and thus the model will be locally identified if both reports are related to true employment and true employment differs between periods.[7]

In the above model, the predicted moments are invariant to replacing every employment event with a non-employment event: replacing $P(S|E)$ with $P(S|E^c)$, replacing $P(E_t)$ with $1 - P(E_t)$, and so on. Thus global identification requires some criterion for selecting between these two locally identified estimates. We assume that the probability of reporting employment is greater when employed than not: $P(S|E) \geq P(S|E^c)$ and $P(A|E) \geq P(A|E^c)$.

III. Matched survey and administrative data

Our identification requires us to estimate the proportion of individuals recorded as employed by both measures, and thus requires our two employment measures to be matched. We use matched survey and administrative data from the IDI. Our primary sample comes from the Survey of Family, Income and Employment (SoFIE), a longitudinal survey collected for 8 annual waves from 2002/2003 until 2009/2010.[8]

[7]Because we are only using pooled cross-sectional moments to identify the model, we estimate average error rates acrosss the population. However, in the empirical analysis, we will estimate models that allow for the error rates to vary across observable characteristics, and also estimate the model separately for demographic subsamples.

[8]The identification presented earlier does not require a longitudinal survey: although multiple time periods must be observed, these can be observed across repeated cross-sections. However observing the

Individuals in the SoFIE sample were matched to administrative data sources in the IDI using name, date of birth and gender. The administrative data is from the Employer Monthly Schedules (EMS), which each employer must file with New Zealand's Inland Revenue tax department. The EMS lists each worker an employer paid earnings to, and thus withheld tax from, in each month.

Our analytical sample consists of the balanced panel of individuals aged 20–64 who completed the full SoFIE survey, could be matched to the IDI data, had no self-employment activity and had no missing employment data or inconsistent annual reference periods over the panel. Table 1 presents summary statistics for this sample, and the excluded sample of working age individuals.

In the first wave of SoFIE, respondents reported employment activity over the 12 months to the end of the previous month. This determined the calendar months for their 'annual reporting periods'. In subsequent waves, respondents were asked about activity since their previous interview, and their employment activity is allocated to their annual reporting period for that wave. The survey measure we use is derived from SoFIE information: we classify a person as employed in a wave if they report any earnings within that wave's annual reporting period. The administrative employment measure classifies an individual as employed if they received any EMS earnings within their SoFIE annual reporting period.

Both of our employment measures have potential error. SoFIE will be subject to recall error – for example, participants may misremember job end dates. The administrative data may be incorrectly matched to SoFIE participants. We have no reason to think the sources of error in the two measures will be correlated.

Table 2 summarises the employment classifications of our sample. Unsurprisingly, the two measures correlate closely: 6.6% of observations show a discrepancy, with observations being slightly more likely to be classified employed in SoFIE and not in the administrative data than vice versa. Discrepancies are greater for women (6.9%) than for men (6.2%), although men are more likely to report employment in SoFIE and not in EMS.

---

cross-time covariance of the sample moments does facilitate optimal minimum-distance estimation.

IV. Results

Table 3 presents an optimal minimum-distance estimation of the model presented in Section II.[9] Estimates in the first four columns were estimated using the full sample. Of these, models (1)–(3) are estimated assuming at least one of the measures is reported without error, while model (4) allows errors in each. Model (1) assumes both measures are reported without error: the false positive rates and false negative rates are restricted to equal zero, and the only free parameters are annual employment rates. Employment rates are estimated to be about 3-4 percentage points lower than the proportions reported as employed in SoFIE or in the administrative data. As suggested by the discrepancies between the two measures, this model fits the data poorly.[10]

Models (2) and (3) each assume one measure is reported without error. These models fit the data more closely, and correspondingly have smaller goodness of fit statistics.[11] Both models precisely estimate a small false negative rate and a large false positive rate in the measure with error.

In column (4), we present the model estimates allowing errors in both reports. The estimated false positive rate in the administrative report is at the zero boundary, while the estimated false negative rate is small, positive and marginally significant.[12] This model results in substantially higher employment rates than those in column (3), which assumed the administrative report was without error.

Figure 1 compares estimated employment rates from models (2), (3) and (4) to

[9] Table 4 contains the sample moments to which the model is fitted and the predicted moments which the model produces. Because some error rates are estimated at the zero-boundary of the probability range, the conditions for the standard asymptotic distribution are invalid, as are the usual full sample bootstrap approach to estimating standard errors (Andrews, 1999). We present subsample bootstrap standard errors based on 100 bootstrap samples of size $N^{7/8}$ (Politis, Romano, & Wolf, 1999).

[10] The model is severely rejected using the formal χ^2 goodness of fit criteria: G.O.F.= 697.4 (with 16 degrees of freedom).

[11] The goodness of fit statistics remain large, and a Sargan overidentification test would easily reject the collective validity of the moment restrictions they require. This is largely due to the first period – when that period is excluded, the goodness of fit statistic for the model displayed in column (4) becomes 18.95, though the other parameters change little. We have also estimated the model allowing for parametric time trends in the error rates. Allowing for linear trends, the model fits the data much more accurately (GoF = 8.2; df = 8). The error rates themselves are little changed, except for the false negative rate in SoFIE which slopes upwards at 0.0089 per year, and the false negative rate for LEED which slopes insignificantly downwards.

[12] When not restricted to be between 0 and 1, the point estimate of the administrative data false positive rate is -0.104. With a 0.098 standard error the hypothesis that it was equal to zero would not be rejected with any confidence.

the annual SoFIE and administrative reported employment rates. Unsurprisingly, the models which assume truth in either SoFIE or in the administrative data estimate employment rates similar to those in the source they assume is true. Despite the model (4) error probability estimates being similar to those in model (3), model (4) estimates higher employment rates, closer to those from model (2).

The estimates in columns (5) and (6) reestimate model (4) separately for men and women. Unsurprisingly, the employment rates for men and women differ; otherwise the estimates resemble those in column (4). For men, the estimated SoFIE false negative rate is actually smaller than that for the administrative data, and a small and statistically-insignificant administrative false positive rate is estimated.

In estimating gender-specific models we are implicitly weakening the assumptions presented in Section II – in particular, we are now requiring independence of errors only after conditioning on gender. This could be justified if gender both predicts false-match rates in the administrative data (perhaps because women are more likely to change their name) and predicts exaggerated survey reports. Our final estimates, in column (7), assume error independence only after conditioning on covariates.[13] Wave-specific logit regressions are estimated with the dependent variable being either SoFIE reporting, administrative data reporting, or reporting in both measures. The moments $P(A_t, S_t|x = \bar{x}), P(A_t|x = \bar{x}), P(S_t|x = \bar{x})$ are calculated by evaluating the estimated logit equations at covariate means. Parameters are then fitted to these predicted moments as before.[14]

Column (7) displays different parameters to those in column (4). The estimated SoFIE false positive rate, for example, is now the probability of reporting employment in SoFIE conditional both on not being employed *and on having covariates equal to* $\bar{x}$. Thus the parameters in column (7) could differ from those in column (4) even if column (4) was correctly specified. Nonetheless, the parameters follow similar contours. When compared to those in column (4), the SoFIE false negative rate is similar, the SoFIE

[13]The covariates are indicator variables for age being less than 30, for age being less than 55, for 5 highest levels of education achieved and for 5 ethnicities, and interactions between those variables and gender.

[14]The covariance matrix used in optimal minimum-distance estimation is bootstrapped with 500 replications, clustering observations across individuals.

false positive rate is insignificantly lower, and both administrative data error rates are insignificantly higher.

In order to allow the error rates to vary across the sample, we have also estimated model (4) on a variety of demographic subsamples. Although these estimates, presented in Table 5, vary across these subsamples, there are no apparent systematic patterns.

Collectively, our estimates suggest that both the administrative data and SoFIE are measured with error. While administrative data errors are rare, with few false negatives and fewer false positives, allowing for erroneous administrative data substantially increases estimated employment rates. There is more error in SoFIE, with false positive rates estimated between 10% and 16%.

We can more formally test the administrative data with the restriction $P(A|E) = 1, P(A|E^c) = 0$. That restriction is rejected for men ($p = 0.046$) and for the conditional moments ($p = 0.005$), but not for women ($p = 0.880$) or for the full sample ($p = 0.500$). We attribute those discrepancies to the low power of overidentification tests, but they may suggest that misclassification in the administrative data varies across subsamples. Similar restrictions on SoFIE are easily rejected in all specifications.

V. Conclusions

The literature on earnings measurement error has generally presumed the absence of error in administrative data. In this study we have shown that assumption to be unjustified. Both measures of employment contain error, and analysis which ignores that error will misrepresent reality. Allowing for error in the administrative reports substantially increases estimated employment rates and, in some subsamples, the restriction that the administrative data lack error can be rejected. Nonetheless, we estimate greater levels of error in the survey data, and in particular between 10% and 16% of those not employed report employment.

References

Abowd, J. M., & Stinson, M. H. (2013). Estimating measurement error in annual job earnings: A comparison of survey and administrative data. *Review of Economics and Statistics*, *95*(5), 1451–1467.

Andrews, D. W. (1999). Estimation when a parameter is on a boundary. *Econometrica*, *67*(6), 1341–1383.

Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. *Handbook of econometrics*, *5*, 3705–3843.

Bound, J., & Krueger, A. B. (1991). The extent of measurement error in longitudinal earnings data: Do two wrongs make a right? *Journal of Labor Economics*, 1–24.

Feng, S., & Hu, Y. (2013). Misclassification errors and the underestimation of the US unemployment rate. *The American Economic Review*, *103*(2), 1054–1070.

Hyslop, D., & Townsend, W. (2016). Earnings dynamics and measurement error in matched survey and administrative data. *Motu Working Paper 16-18*.

Kapteyn, A., & Ypma, J. Y. (2007). Measurement error and misclassification: A comparison of survey and administrative data. *Journal of Labor Economics*, *25*(3), 513–551.

Keane, M. P., & Sauer, R. M. (2009). Classification error in dynamic discrete choice models: Implications for female labor supply behavior. *Econometrica*, *77*(3), 975–991.

Pischke, J.-S. (1995). Measurement error and earnings dynamics: Some estimates from the PSID validation study. *Journal of Business & Economic Statistics*, *13*(3), 305–314.

Politis, D. N., Romano, J. P., & Wolf, M. (1999). *Subsampling*. New York: Springer Verlag.

Poterba, J. M., & Summers, L. H. (1986). Reporting errors and labor market dynamics. *Econometrica*, 1319–1338.

Poterba, J. M., & Summers, L. H. (1995). Unemployment benefits and labor market transitions: A multinomial logit model with errors in classification. *The Review of Economics and Statistics*, 207–216.

Figure 1: *Estimated and sample employment rates by wave*

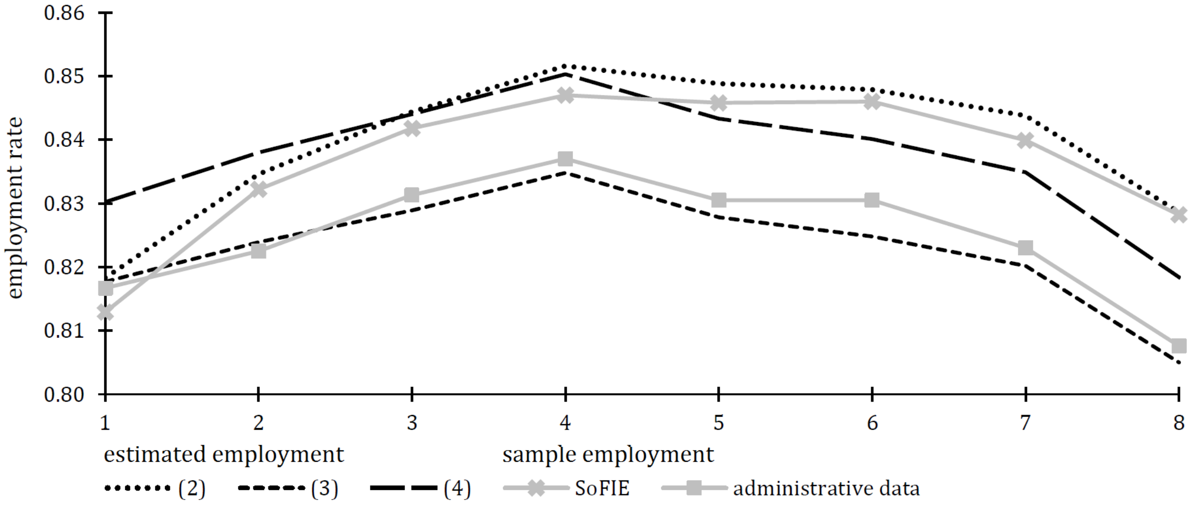

Table 1: *Summary statistics*

	Excluded sample (1)	Analysis sample (2)
SoFIE characteristics:		
Age	40.62	43.04
	(14.0)	(10.4)
Female	0.48	0.59
	(.50)	(.49)
European	0.74	0.77
	(.44)	(.42)
Degree-educated	0.18	0.19
	(.38)	(.39)
SoFIE reported:		
Employed	0.62	0.84
	(.49)	(.37)
Earnings ($)	20,938	35,010
	(30,051)	(37,502)
Self-employed	0.30	0
	(.46)	
EMS admin reported:		
Employed	0.62	0.82
	(.48)	(.38)
Earnings ($)	20,462	34,882
	(29,320)	(34,486)
No. observations	63,339	56,832
No. individuals	18,438	7,104

Notes: Standard deviations in parentheses. Counts are randomly rounded to base 3. The analysis sample is the balanced panel of matched individuals aged 20–64. The excluded sample consists of SoFIE individuals not appearing in all eight waves, as well as 3% of individuals unmatched in the IDI, 10% with self-employment, and 54 with missing employment or changing annual reference periods.

Table 2: *Measurement discrepancies*

	EMS administrative reports		
	not employed	employed	all
	A: All individuals		
SoFIE reports:			
not employed	0.136	0.027	0.163
employed	0.039	0.798	0.837
all	0.175	0.825	1
No. observations	56,832		
No. individuals	7,104		
	B: Men		
SoFIE reports:			
not employed	0.076	0.018	0.094
employed	0.044	0.862	0.906
all	0.120	0.880	1
No. observations	23,034		
No. individuals	2,880		
	C: Women		
SoFIE reports:			
not employed	0.177	0.033	0.210
employed	0.036	0.754	0.790
all	0.213	0.787	1
No. observations	33,792		
No. individuals	4,224		

Notes: Counts are randomly rounded to base 3 and so subtotals do not add to totals.

Table 3: *Minimum-distance estimates*

	All individuals				Men	Women	Repr. $\bar{x}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1 - $P(A\|E)$	—	0.045	—	0.016	0.032	0.011	0.039
		(.003)		(.012)	(.006)	(.013)	(.000)
$P(A\|E^c)$	—	0.149	—	0.000	0.028	0.000	0.009
		(.011)		(.044)	(.026)	(.047)	(.013)
1 - $P(S\|E)$	—	—	0.030	0.030	0.017	0.039	0.028
			(.002)	(.009)	(.004)	(.013)	(.002)
$P(S\|E^c)$	—	—	0.213	0.156	0.157	0.130	0.099
			(.011)	(.050)	(.054)	(.044)	(.003)
$P(E_1)$	0.773	0.818	0.818	0.830	0.903	0.783	0.866
	(.005)	(.005)	(.005)	(.008)	(.009)	(.012)	(.003)
$P(E_2)$	0.784	0.835	0.824	0.838	0.909	0.793	0.876
	(.006)	(.005)	(.005)	(.007)	(.009)	(.010)	(.003)
$P(E_3)$	0.792	0.844	0.829	0.844	0.911	0.804	0.883
	(.006)	(.005)	(.005)	(.008)	(.009)	(.010)	(.002)
$P(E_4)$	0.799	0.852	0.835	0.850	0.914	0.811	0.892
	(.006)	(.005)	(.005)	(.007)	(.009)	(.010)	(.002)
$P(E_5)$	0.792	0.849	0.828	0.843	0.914	0.801	0.884
	(.006)	(.005)	(.005)	(.008)	(.009)	(.010)	(.002)
$P(E_6)$	0.790	0.848	0.825	0.840	0.908	0.799	0.881
	(.006)	(.005)	(.005)	(.009)	(.009)	(.010)	(.002)
$P(E_7)$	0.786	0.844	0.820	0.835	0.904	0.791	0.874
	(.006)	(.005)	(.005)	(.008)	(.009)	(.009)	(.002)
$P(E_8)$	0.770	0.829	0.805	0.818	0.885	0.777	0.856
	(.006)	(.006)	(.005)	(.008)	(.009)	(.009)	(.002)
No. individuals	7,104	7,104	7,104	7,104	2,880	4,224	7,104
No. observations	56,832	56,832	56,832	56,832	23,034	33,792	56,832
GoF	697.4	109.2	90.6	89.2	26.8	75.4	78.3
(df)	(16)	(14)	(14)	(12)	(12)	(12)	(12)

Notes: Estimates from optimal minimum-distance estimation of employment rates and misclassification error rates. Standard errors in parentheses, estimated using the subsample bootstrap method with 100 bootstrap replications of bootstrap samples of size $N^{7/8}$. Model (1) estimated assuming no errors in SoFIE and EMS admin reports, model (2) assumes no errors in SoFIE reports, model (3) assumes no errors in EMS admin reports, models (4) – (7) allow errors in both reports. The GoF statistics for the models are χ^2_{df} distributed under the null hypothesis that the model provides an adequate fit to the sample moments.

Table 4: *Sample and predicted moments*

	All individuals					Men		Women		Repr. $\bar{x}$	
	sample	predicted using model				sample	model	sample	model	sample	model
		(1)	(2)	(3)	(4)		(5)		(6)		(7)
$P(A_1, S_1)$	0.774	0.773	0.781	0.793	0.792	0.849	0.859	0.722	0.745	0.791	0.808
$P(A_2, S_2)$	0.793	0.784	0.797	0.799	0.800	0.862	0.865	0.746	0.754	0.812	0.818
$P(A_3, S_3)$	0.803	0.792	0.806	0.804	0.805	0.866	0.867	0.759	0.765	0.819	0.824
$P(A_4, S_4)$	0.810	0.799	0.813	0.810	0.812	0.870	0.870	0.770	0.772	0.827	0.832
$P(A_5, S_5)$	0.808	0.792	0.811	0.803	0.805	0.871	0.870	0.766	0.762	0.826	0.825
$P(A_6, S_6)$	0.807	0.790	0.810	0.800	0.802	0.869	0.865	0.764	0.760	0.824	0.822
$P(A_7, S_7)$	0.801	0.786	0.806	0.796	0.797	0.866	0.861	0.757	0.752	0.818	0.816
$P(A_8, S_8)$	0.787	0.770	0.791	0.781	0.781	0.846	0.842	0.747	0.739	0.802	0.799
$P(A_1)$	0.817	0.773	0.808	0.818	0.817	0.878	0.876	0.758	0.775	0.834	0.833
$P(A_2)$	0.823	0.784	0.822	0.824	0.825	0.882	0.882	0.783	0.784	0.841	0.842
$P(A_3)$	0.831	0.792	0.829	0.829	0.831	0.885	0.884	0.795	0.796	0.848	0.849
$P(A_4)$	0.837	0.799	0.835	0.835	0.837	0.887	0.887	0.803	0.803	0.854	0.857
$P(A_5)$	0.831	0.792	0.833	0.828	0.830	0.885	0.887	0.802	0.792	0.848	0.850
$P(A_6)$	0.831	0.790	0.832	0.825	0.827	0.884	0.882	0.800	0.791	0.847	0.847
$P(A_7)$	0.823	0.786	0.829	0.820	0.822	0.881	0.878	0.792	0.782	0.840	0.840
$P(A_8)$	0.808	0.770	0.817	0.805	0.805	0.862	0.859	0.785	0.769	0.824	0.823
$P(S_1)$	0.813	0.773	0.818	0.832	0.832	0.894	0.903	0.775	0.781	0.838	0.855
$P(S_2)$	0.832	0.784	0.835	0.837	0.838	0.905	0.908	0.782	0.789	0.859	0.863
$P(S_3)$	0.842	0.792	0.844	0.841	0.843	0.910	0.909	0.795	0.799	0.866	0.869
$P(S_4)$	0.847	0.799	0.852	0.845	0.848	0.912	0.912	0.803	0.804	0.870	0.877
$P(S_5)$	0.846	0.792	0.849	0.840	0.842	0.910	0.912	0.793	0.796	0.869	0.871
$P(S_6)$	0.846	0.790	0.848	0.838	0.840	0.914	0.907	0.794	0.795	0.873	0.868
$P(S_7)$	0.840	0.786	0.844	0.834	0.835	0.911	0.904	0.784	0.787	0.866	0.862
$P(S_8)$	0.828	0.770	0.829	0.823	0.822	0.892	0.888	0.771	0.776	0.850	0.846

Sample moments and predicted moments corresponding to the models displayed in Table 3. Sample moments are estimated with sample proportions, except for those for the representative $\bar{x}$ which are estimated with a logit model evaluated at covariate means.

Table 5: *Model estimates from subsamples*

	Ethnicity		Age			Education		Women, partnered		
	Euro	Other	Young	Prime	Old	School	Tertiary	Always	Never	Some
$1 - P(A\|E)$	0.006	0.028	0.000	0.000	0.039	0.007	0.018	0.000	0.048	0.000
	(.009)	(.016)	(.003)	(.015)	(.016)	(.008)	(.014)	(.014)	(.172)	(.009)
$P(A\|E^c)$	0.000	0.156	0.000	0.139	0.065	0.000	0.188	0.147	0.110	0.000
	(.027)	(.032)	(.051)	(.051)	(.031)	(.023)	(.072)	(.048)	(.134)	(.046)
$1 - P(S\|E)$	0.026	0.000	0.045	0.000	0.011	0.034	0.000	0.000	0.000	0.042
	(.005)	(.007)	(.009)	(.008)	(.063)	(.008)	(.008)	(.012)	(.007)	(.069)
$P(S\|E^c)$	0.210	0.104	0.145	0.245	0.070	0.147	0.183	0.176	0.000	0.182
	(.048)	(.045)	(.023)	(.068)	(.070)	(.029)	(.073)	(.054)	(.076)	(.106)
$P(E_1)$	0.848	0.698	0.842	0.793	0.797	0.768	0.835	0.744	0.712	0.810
	(.007)	(.016)	(.016)	(.011)	(.021)	(.010)	(.010)	(.016)	(.105)	(.069)
$P(E_2)$	0.852	0.724	0.849	0.804	0.802	0.773	0.851	0.754	0.736	0.821
	(.007)	(.017)	(.013)	(.012)	(.051)	(.010)	(.009)	(.017)	(.098)	(.044)
$P(E_3)$	0.852	0.751	0.871	0.811	0.798	0.782	0.859	0.761	0.754	0.834
	(.007)	(.018)	(.011)	(.013)	(.051)	(.010)	(.011)	(.017)	(.110)	(.036)
$P(E_4)$	0.857	0.766	0.874	0.822	0.794	0.793	0.863	0.775	0.753	0.840
	(.008)	(.018)	(.011)	(.014)	(.022)	(.010)	(.013)	(.015)	(.115)	(.028)
$P(E_5)$	0.848	0.776	0.852	0.817	0.788	0.782	0.860	0.763	0.762	0.823
	(.008)	(.021)	(.011)	(.016)	(.051)	(.010)	(.014)	(.016)	(.047)	(.026)
$P(E_6)$	0.846	0.770	0.856	0.816	0.773	0.779	0.858	0.759	0.770	0.814
	(.009)	(.017)	(.013)	(.017)	(.050)	(.010)	(.014)	(.018)	(.093)	(.088)
$P(E_7)$	0.842	0.764	0.842	0.812	0.757	0.775	0.851	0.752	0.765	0.802
	(.008)	(.015)	(.012)	(.017)	(.023)	(.009)	(.013)	(.015)	(.108)	(.033)
$P(E_8)$	0.827	0.748	0.831	0.798	0.729	0.759	0.832	0.739	0.762	0.783
	(.008)	(.017)	(.011)	(.018)	(.046)	(.010)	(.014)	(.016)	(.083)	(.054)
No. individuals	5,454	1,650	1,332	4,326	1,446	3,018	4,083	2,118	669	1,332
No. observations	43,632	13,200	10,656	34,608	11,568	24,144	32,664	16,944	5,352	10,656
GoF	58.2	30.9	35.1	73.9	21.3	46.6	64.6	25.7	16.7	53.0

Standard errors in parentheses, estimated using the subsample bootstrap method with 100 bootstrap replications of bootstrap samples of size $N^{7/8}$. All GoF statistics have 12 degrees of freedom.

Recent Motu Working Papers

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

16-18 Hyslop, Dean and Wilbur Townsend. 2016. “Earnings Dynamics and Measurement Error in Matched Survey and Administrative Data.”

16-17 Hollis, Michele, Cecile de Klein, Dave Frame, Mike Harvey, Martin Manning, Andy Reisinger, Suzi Kerr, and Anna Robinson. 2016. “Cows, Sheep and Science: A Scientific Perspective on Biological Emissions from Agriculture”

16-16 Kerr, Suzi. 2016. “Agricultural Emissions Mitigation in New Zealand: Answers to Questions from the Parliamentary Commissioner for the Environment.”

16-15 Timar, Levente. 2016. “Yield to Change: Modelling the Land-use Response to Climate Driven Changes in Pasture Production.”

16-14 Chappell, Nathan and Adam B Jaffe. 2016. “Intangible Investment and Firm Performance.”

16-13 Kerr, Suzi and Judd Ormsby. 2016. “The New Zealand Emissions Trading Scheme de-link from Kyoto: impacts on banking and prices.”

16-12 Carver, Tom and Arthur Grimes. 2016. “Income or consumption: which better predicts subjective wellbeing?”

16-11 Apatov, Eyal and Arthur Grimes. 2016. “Higher education institutions and regional growth: The case of New Zealand.

16-10 Chappell, Nathan and Isabelle Sin. 2016. “The Effect of Trial Periods in Employment on Firm Hiring Behaviour” (also known as Treasury Working Paper 16/03)

16-09 Timar, Levente. 2016. “Does money grow on trees? Mitigation under climate policy in a heterogeneous sheep-beef sector”

16-08 Jaffe, Adam, Trinh Le and Nathan Chappell. 2016. “Productivity distribution and drivers of productivity growth in the construction industry.”

16-07 Fabling, Richard, Arthur Grimes and Levente Timar. 2016. “Labour market dynamics following a regional disaster.”

16-06 Leining, Catherine and Suzi Kerr. 2016. “Lessons Learned from the New Zealand Emissions Trading Scheme.”

16-05 Grimes, Arthur, Judd Omsby, Anna Robinson, Siu Yuat Wong. 2016. “Subjective wellbeing impacts of national and subnational fiscal policies.”

16-04 Ryan Greenaway-McGrevy, Arthur Grimes, Mark Holmes. 2016. “Two Countries, Sixteen Cities, Five Thousand Kilometres: How Many Housing Markets?”

16-03 Fabling, Richard and Lynda Sanderson. 2016. “A Rough Guide to New Zealand’s Longitudinal Business Database (2nd edition).”

16-02 MacCulloch, Robert. 2016 “Can “happiness data” help evaluate economic policies?”

16-01 Gørgens, Tue and Dean Hyslop. 2016. “The specification of dynamic discrete-time two-state panel data models”

15-20 Maré David C., Ruth M. Pinkerton and Jacques Poot. 2015. “Residential Assimilation of Immigrants: A Cohort Approach.”

15-19 Timar, Levente, Arthur Grimes and Richard Fabling. 2015. “Before a Fall: Impacts of Earthquake Regulation and Building Codes on the Commercial Market”

15-18 Maré David C., Dean R. Hyslop and Richard Fabling. 2015. “Firm Productivity Growth and Skill.”

15-17 Fabling, Richard and David C. Maré. 2015. “Addressing the absence of hours information in linked employer-employee data.”

Motu