

Session Number: **6B**
Session Title: **Self-Employment and Inequality**
Session Chair: **Peter Saunders**

*Paper Prepared for the 29th General Conference of
The International Association for Research in Income and Wealth*

Joensuu, Finland, August 20 – 26, 2006

Self-Employment and the Earnings Distribution:
Have Globalization and European Integration caused Convergence?

Dennis H. Sullivan, Miami University, Oxford, Ohio, USA
(sullivdh@muohio.edu)

Timothy M. Smeeding, Syracuse University, Syracuse, NY, USA
and Luxembourg Income Study
(tmsmeed@maxwell.syr.edu)

This paper is posted on the following website: <http://www.iariw.org>

I. Introduction.

How much do countries differ in the extent to which workers are self-employed? Are the consequences for the income distribution similar in every country, or do they differ in systematic ways? There is really no disputing that self-employment is an important labor market phenomenon that ought to have distributive consequences. The most recent OECD data (for 2004) on the proportion of the labor force that is self-employed show an “OECD total” of 17.4%, with a wide range, from less than 8% in Norway and the US to more than 30% in Korea, Mexico and Turkey.

Cross-national studies of self-employment were sparse in the 1990’s. For example, the survey by Le (1999) contains not a single example. The working paper by Acs, Audretsch and Evans (1994) is the most commonly cited paper from the period. Greater attention has been paid to self-employment in the last several years, including articles by Robson and Wren (1999), Blanchflower (2000), Parker and Robson (2004) and Torrini (2005). Key results of these recent articles are summarized in the next section.

With the partial exception of Blanchflower (2000), the recent papers focus almost exclusively on the *causes* of cross-national differences in the prevalence of self-employment, and treat self-employment as a decision made by workers, rather than households. The research has little to say about the *consequences* of cross-national differences, particularly the income distribution consequences. While the OECD provides a one-stop shop for much of the data required to study the causes of self-employment among the labor force, the Luxembourg Income Study (LIS) can be used to supplement the OECD to focus on household measures of self-employment and the distributive effects of self-employment.

In a working paper written almost a decade ago (Sullivan and Smeeding, 1997) we found that self-employment plays a major role in altering the market income distribution in every one of the nineteen countries we studied. In particular, we found that households reliant exclusively on self-employment income tended to clump in the upper and lower tails of the market income distribution, that households that combine self-employment with employment earnings were concentrated in the upper tail of the distribution, and that these broad distributive patterns were to be found in almost every country in the dataset.

This paper is a sequel that asks three basic questions:

- 1) Are the patterns of household self-employment and its distributional consequences that we found in Wave III (around 1990) of LIS still to be found in the larger and more recent sample of countries in Wave V (around 2000)?
- 2) Has there been noticeable cross-national convergence during the 1990’s, particularly in Europe (perhaps due to the Single Market and EU enlargement)?
- 3) Do the variables that seem to “explain” cross-national differences in the prevalence of self-employment also “explain” cross-national differences in the distributive consequences? If so, which are the most important?

II. Previous Research About Self-Employment in OECD Countries.

It is often the case that data designs determine research designs, and the study of self-employment in OECD countries is a prize example. The OECD maintains a time series of the percentage of the civilian labor force that is self-employed and a raft of potential regressors to put on the right side of a reduced-form regression. The seminal example is Acs, Audretsch and Evans (1994), a working paper that established the basic stylized fact that the prevalence of self-employment in a country is negatively correlated with the country's level of development, and that this stylized fact seems robust when other controls are included.

The first important published paper using panel data techniques to explore the self-employment rate was Robson and Wren (1999). The focus of their paper is upon the effect of taxes-- a natural focus, since one of the obvious differences across national borders is differences in tax regimes. They find that high marginal tax rates tend to discourage self-employment and that high (and rising) average tax burdens tend to increase self-employment rates. Other significant controls include the level of per capita GNP (negative effect on self-employment rate, as expected), growth of unemployment, growth of manufacturing employment and level of service employment (all negative effects). Two subsequent papers that built upon Robson and Wren (1999) were Fölster (2002) and Robson (2003). Fölster verifies the importance of taxes, while Robson finds that careful analysis undermines the view that self-employment is heavily influenced by employer protection legislation.

The latest contributions to the study of OECD self-employment rates are Parker and Robson (2004) and Torrini (2005). Parker and Robson show that six of the twelve countries they survey have had increasing self-employment rates, three have had decreasing self-employment rates, and three have had roughly constant self-employment rates. Since the resulting panel data have substantial variation based on different trends as well as different levels, the authors are able to deploy an alternative dynamic panel technique ("Fully Modified OLS"—see Pedroni (2000)), and they discover that their results are quite sensitive to the technique employed. For example, the received conclusion that self-employment rates fall with real per capita GDP is sensitive to technique, as is the sign on the average tax rate. The other variables that have significant effects are the female labor force participation rate (negative) and rate at which unemployment benefits replace earnings (negative when significant).

Torrini (2005) follows a similar path, though he uses a different dynamic panel technique ("Dynamic OLS"—see Mark and Sul (2003)). His results are less sensitive to the technique chosen, with a negative effect of per capita GDP, a negative effect of public sector size, a positive effect of the unemployment rate, and possibly a negative effect of the tax "wedge." He verifies Robson's (2003) conclusion that employment protection legislation is unimportant, and he introduces two additional variables: a "corruption index" (which he interprets as evidence that tax laws may not be strictly enforced) and a measure of product market regulation. Although the introduction of new institutional variables is useful, because they define other measures of the role of national borders, the

particular variables do not work out very well, apparently because of high correlations with other variables.

Blanchflower's (2000) paper begins on a track similar to the others, but then veers off in a rather different direction. After starting off with an analysis somewhat similar to that of Robson and Wren (1999) (the authors do not seem to have been aware of one another's research), Blanchflower proceeds to the analysis of concatenated microdata for 19 countries collected over various lengths of time (in some cases over twenty years). He finds that the probability of self-employment rises with age and that the effect of the unemployment rate is country-specific (and often negligible).

III. Self-Employment Income Data: Problems and Prospects

We feel compelled to begin with the following truth-in-packaging disclaimer: survey data on self-employment income are not particularly accurate, and the degree of inaccuracy varies across countries. In a careful study of this matter, Atkinson, Rainwater and Smeeding (1995) provide evidence that "Self-employment income reporting differs substantially across surveys" (p. 36); the accompanying table suggests that, compared to national income aggregates, self-employment income is generally much more severely underreported than wage and salary income, and that the degree of underreporting varies across countries.

We have three reasons for embarking on the analysis in spite of our concerns about data quality. The first is that data quality matters less for some analyses than others. For example, as Gottschalk and Smeeding (1997) emphasize, systematic inaccuracies that arise from factors like cross-national differences in survey methodologies will affect relative *trends* of measured income much less than relative *levels*, so the data problems should be less of a concern when we ask whether self-employment metrics have *converged*. Second, an analysis that defines self-employment as "receipt of self-employment income" may shed a different light on the phenomenon itself. For example, our previous paper (1997) showed that the stereotype that self-employment is dominated by family firms that fully employ the household's workers is simply wrong: in most of the LIS countries, the majority of households with self-employment income also have wage and salary income. Third, and most important, we want to move the analysis away from the usual preoccupation with *why* more workers choose self-employment in some countries than in others, and instead ask, "*So what?*" For example, another conclusion from our previous analysis was that households with self-employment income tend to be concentrated in the tails of the market income distribution in almost every LIS country.

IV. Self-Employment Income in the Luxembourg Income Study

Every one of the thirty countries in the Luxembourg Income Study has at least one survey that includes a measure of household self-employment income. In fact, the Wave V (around the year 2000) datasets contain a variable that measures self-employment income at the level of the *individual*, though we stick to the *household* measure, partly because our emphasis is upon the household income distribution and partly to maintain comparability over time. In addition to the underreporting problem discussed above, the principal dilemma with these data is what to do about households that report losses from self-employment. Some surveys simply forbid them to report these losses as negative income, while other countries employ a bottom-code that is less than zero. Our response is twofold. First, we bottom-code self-employment income to 1 unit of local currency, so that non-positive reports are treated in a comparable way. Second, we employ measures that are relatively insensitive to the resolution of the dilemma: a household that reports negative self-employment income is still a household that reports such income, and if self-employment is its only source of earnings, it will be in the bottom quintile whether the report is recorded as negative or zero.

Turning to distributional questions, we need to define the exact measure of “income,” the distribution of which we propose to study. That measure will be “household non-farm earnings,” that is, the sum of the household’s wage and salary income and its non-farm self-employment income (bottom-coded to 1 unit of local currency). Households that have neither form of earnings (for example, households completely reliant on farm income or pensions or transfer payments) are simply excluded from the sample. Another problem is that some countries only report wage and salary incomes net of taxes. Other things equal, we would expect the self-employed to have higher relative earnings in those countries, and accordingly we have noted the identities of those countries in the tables by showing the country names in *italic type*.

V. Results: Proportion of Households with Self-Employment Income

The first question we want to ask is: what proportion of households have non-farm self-employment income (SEI) as their entire source of earnings? The results for Wave III (around 1990) and Wave V (around 2000) are shown in Tables 1 and 2, arranged as league tables from the highest to the lowest percentage with SEI only. Given the well-known correlation with per capita GDP, it is not surprising that Taiwan and Mexico are high in the tables. The presence of Italy, Spain and Greece near the top conforms with stereotypes of their economies as having a large informal sector of self-employed households. In Wave III, near the beginning of the transition period, the formerly Communist countries are all low in the table, though by Wave V, a decade later, Russia and Poland are high in the table, as one might expect.

One of the surprises from looking at self-employment through the lens of income is how many households report both self-employment income (SEI) and wage and salary income (WSI). In fact, in a majority of countries—13 out of 21 in Wave III, 16 out of 24 in Wave

V--, these “combiners” outnumber the households with SEI only. As indicated by the correlation coefficients on the tables, the proportion of a nation’s households that combine SEI with a smaller amount of WSI is highly correlated with the proportion that report SEI only. Those who combine WSI with a smaller amount of SEI are rather different: this group is especially prevalent in the North American and Nordic economies.¹

VI. Results: Percentages in Bottom and Top Quintiles.

The next question we want to ask is whether certain groups are overrepresented or underrepresented in the bottom quintile of the overall distribution of non-farm earnings. The results are shown in Tables 3 and 4. The basic result is dramatic: **households with SEI only are overrepresented in the bottom quintile, and households that combine SEI with WSI are underrepresented in the bottom quintile**, in practically every country in both years.

Obviously this result can be explained in part by systematic underreporting of SEI, but ought not underreporting also affect the households that have SEI>WSI? Speaking generally, the greatest overrepresentation among the SEI only (and the least underrepresentation among the SEI>WSI) is to be found in the Nordic countries, Switzerland, Canada, the US and the UK—countries that tend to conduct relatively careful surveys. The countries with the least overrepresentation are, generally speaking, those in which self-employment income is most prevalent.

The corresponding results for the top quintile are given as Tables 5 and 6. Again there is a clear result: **households with SEI only are underrepresented in the top quintile, and households that combine SEI with WSI are overrepresented in the top quintile**, in practically every country in both years. Note that underreporting of SEI cannot be the culprit in this case. Indeed, in most countries those with SEI>WSI are *more* overrepresented than those with WSI>SEI. In general, the Nordic countries are low on this particular league table, and it is interesting that Canada, the UK and the US all have overrepresentation statistics that are well below the cross-national averages.

The countries that are high in Tables 1 and 2 are generally low in Tables 3 and 4 and high in Tables 5 and 6; the countries that are low in Tables 1 and 2 are generally high in Table 3 and low in Tables 5 and 6. There are some exceptions², but it seems to be a valid generalization that **in countries in which a larger proportion of households earn self-employment income, those households are less overrepresented in the bottom quintile and more overrepresented in the top quintile.**

¹ This result may be distorted somewhat by the fact that the Nordic economies, Canada and the US all report gross wage and salary income, making WSI>SEI somewhat more likely. There is not, however, a close association overall between the relative sizes of the two groups of “combiners” and whether wage and salary income is measured net of taxes.

² The rank correlation between Tables 1 and 3 =-0.53; between Tables 1 and 5 =0.14 ; between Tables 2 and 4 =-0.35; and between Tables 2 and 6 =0.20.

VII. Have Self-Employment Metrics Converged?

Before answering the question in the section title, we must first answer, “Which self-employment metrics are we talking about?” We have selected four of them: 1) the percentage of households with SEI only; 2) the percentage of households with some SEI; 3) the percentage of the SEI-only households in the 1st (bottom) quintile; and 4) the percentage of the SEI-only households in the top quintile.

A second preliminary question is, “How is ‘convergence’ measured?” We have selected two measures: 1) the change in the coefficient of variation of the metric between Wave III and Wave V; and 2) the coefficient from regressing the change in the metric on its initial (Wave III) value. In both instances, then, “convergence” will be interpreted as a negative value.

The third (and last!) preliminary question is, “Convergence across what geographical area?” We first calculate the convergence metrics for the thirteen countries for which they can be calculated for both Wave III and Wave V: Canada, Finland, Germany, Israel, Italy, Mexico, Netherlands, Norway, Spain, Sweden, Switzerland, UK, and US. We then restrict the sample, first to the countries in either Europe or NAFTA, then to those in Europe only (including, of course, two non-EU countries, Norway and Switzerland). One might expect that convergence would be more obvious in the narrower groupings.

The results are displayed in Table 7. The simple answer to the question, “Have self-employment metrics converged?” is, “Not really.” Fewer than half of the values in Table 7 are negative, and only one is statistically significantly negative. Unfortunately, the reliability of these calculations is suspect, because four countries, all of them European, changed surveys between Wave III and Wave V: Luxembourg, Spain, Switzerland and the United Kingdom. In at least two of these countries, Luxembourg and Switzerland, the self-employment metrics changed to a degree that cannot plausibly be explained as a real economic change. It is obvious that omitting Luxembourg would cause the metrics to change in the direction of greater convergence (or lesser divergence), though it turns out that omitting all four countries does not alter the results very much at all.

In sum, the evidence for convergence of self-employment metrics is both slender and fragile, depending on both the method for measuring convergence and the selection of countries to which it is applied. While the prevalence of self-employment rose in some countries and fell in others (the latter group was larger), both the *relative* prevalence and the *relative* income distribution effects vary more across countries than over time.

VIII. What Correlates with Cross-National Variation?

The essential innovation of this paper is that it uses the Luxembourg Income Study to develop alternative variables for analysis. The question we want to ask in this section is how the analysis of these new variables differs from the analyses of the OECD self-employment data analyzed by Robson and Wren (1999), Blanchflower (2000), Parker and Robson (2004) and Torrini (2005). We want to emphasize that we are strictly looking at a single cross-section, whereas these previous authors exploited the panel properties of the OECD series to one degree or another. Our goal is to determine whether variables similar to those employed by these previous authors seem to account for the cross-section variation of our variables.

We have selected the same four analysis variables as the last section: 1) the percentage of households with SEI only; 2) the percentage of households with some SEI; 3) the percentage of the SEI only households in the 1st (bottom) quintile; and 4) the percentage of the SEI only households in the top quintile. We have selected eight covariates: 1) real GDP per capita; 2) Service industries as a percentage of value added; 3) unemployment rate; 4) percentage of the labor force with a “tertiary” degree; 4) Taxes as a % of GDP; 5) the Transparency International “Corruption Index,”³ as suggested by Torrini (2005); 6) an index of “Barriers to Entrepreneurship” devised by the OECD; and 7) a “Capital Access Index”⁴ devised by the Milken Institute. All of these data are available from the OECD website except #5) and #7), the origins of which are provided in the footnotes. We have matched these data with LIS data for as many countries as possible, using the Wave V data when available, and the Wave IV data otherwise, providing a dataset of 23 countries. When the covariates are available for multiple years, we have chosen a year as close to the survey year as possible.

Table 8 displays a correlation matrix. The top panel treats each nation as a single observation, regardless of the nation’s size. In the lower panel the calculations are weighted by population. We see first of all that the correlation between “% SEI only” and “% Some SEI” is very close,⁵ even though, as we discussed above, households combining SEI with WSI are often a majority of those with self-employment income.

The correlation between “% SEI only” and the proportion of the SEI-only group in the bottom quintile is *negative* for both the weighted and unweighted calculations: a higher proportion of households (only) self-employed is associated with a smaller proportion of (only) self-employed households in the bottom quintile. Though one hesitates to advance a causal chain to explain this correlation, it seems reasonable that reliance upon self-employment will be a more attractive when there is less evidence that being reliant on self-employment is a lottery ticket with a high probability of being assigned to the bottom of the income distribution. The correlation between “% SEI only” and the proportion of the SEI-only group in the top quintile is virtually zero for the unweighted calculation, but

³ <<http://www1.transparency.org/cpi/2000/cpi2000.html#cpi>>

⁴ See Yago, et.al., 2000.

⁵ The (unweighted) correlation with the OECD self-employment rate is 0.901 for “% SEI only” and 0.826 for “% Some SEI”.