The German Taxpayer-Panel

By Susan Kriete-Dodds and Daniel Vorgrimler

The use of panel data has become increasingly popular in socioeconomic research. This is due to the new analysis possibilities presented by the additional dimension of time, new analysis techniques and the increasing availability of panel data (e.g. the panel of the micro census). Particularly popular are panels based on income tax data (so-called taxpayer-panels). Via "new tax responsiveness" (NTR), a research field has been established which analyses tax-related questions through the use of panel data (e.g. the impact of a tax reform).

An early German foray into this field was by Gottfried / Schellhorn (2003). This work was based on a panel of income taxpayer-panels from Baden-Württemberg with around 50,000 income files from the years 1988 to 1991. Other panels from German income tax data are not known. The absence of panel data is the reason for the little impact these data have had so far on economic discussion within Germany.

Using income tax data, the Federal Statistical Office Germany has succeeded in developing a panel for the study of German taxpayers. The major challenge in the development of this panel was the fact that the identifier for an individual (the tax number) can change during an individual's life (e.g. by moving to another state, marriage or divorce). To tackle this problem, a three-step matching algorithm was developed, so that representative and current income tax panel data are now available in Germany for the first time.

The purpose of this paper is to present this matching algorithm and to demonstrate a number of potential analyses than can be done with the data.

1. Data

The taxpayer-panel is based on the annual income tax statistics, which contain the complete income tax forms of around 28 million taxpayers. Even without being linked over time these data are already used in politics as well as in economic research since they can show the structural effects of income tax as well as the distribution of income and tax burden.

These data are available at the Federal Statistical Office Germany from 2001, thus the currently-available data covers the years 2001 and 2002 and about 95 % of the data for 2003. These three years form the basis of the tax-payer-panel. The time-lag is caused by the amount of time German taxpayers have to hand in their income tax form, which is about 3 years. For the year 2004 only 60 % of the data are available, therefore, it will be only used in parts of the matching process.

The high number of variables involved is a result of the complexity of German income tax law, and makes it possible for users to analyse not only tax properties, but also social ones; such as how income changes after a taxpayer relocates.

Taxpayers are households rather than individuals, since married couples generally hand in one tax claim even if they both have an income.

2. Matching the Data

In principle, the data of the annual income tax statistics could be linked by using the income tax numbers. The problem is that the tax number can change over time, for example after marriage, divorce, moving from one federal state to another, or when the income type[1] changes. Using only the tax number would lead to an enhanced panel mortality, which would include income tax-payers that could not be matched because of missing indicators as well as tax-payers that left the population or entered it. While the first subset should be minimised, the second subset is useful for addressing particular points of inter-est. Since both cases appear at the same time, the two subsets need to be dif-ferentiated clearly. Therefore, a three-step linkage procedure was used. These steps are shown in figure 1 and will be explained in the following.

2.1 Linkage with the Tax Number

While it cannot be used for Germany as a whole, the income tax number can be used as a direct identifier on the basis of the federal states, which only changes, as mentioned above, for certain reasons such as marriage, divorce, etc.

By using the tax number, about 20.6 million taxpayers could be linked for the years 2001 to 2003. Apart from these taxpayers, there were about 3.1 mil-lion taxpayers that appeared in the panel only for the years 2001 and 2002, plus another 1.9 million that submitted no tax declaration in 2001 but did so in 2002 and 2003.

---

[1] German tax law defines seven types of incomes, which are determined by the in-come source (land produce, regular salary, dividends, etc.).
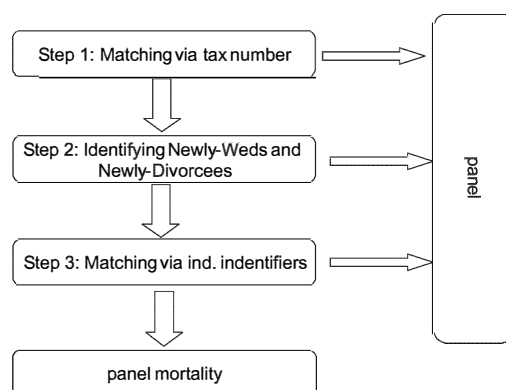
Figure 1

Upon changing a tax number, the tax offices should keep the old tax number on file. In practice this is not always possible, such as when a taxpayer moves to another federal state.

Nevertheless, a further 63,000 taxpayers could be identified who received a new tax-number in 2002 but the tax office tracked these people using their old number. If the old tax number was not separately provided, a portion of the taxpayers who changed their tax numbers could be linked via the indirect factors detailed in section 2.3. For the period 2002 – 2003, approximately 294,000 taxpayers could be linked via their old tax numbers (see Table 1).

Table 1

The German Taxpayer-Panel 2001 – 2003 by Matching Steps

		Years	
Matching Step		2001 / 2002	2002 / 2003
		In 1000	
Step 1	Income Tax Number	24331.3	23010.5
	Old Tax Number	62.5	293.9
	Irregular Employment History	313.4	276.5
Step 2	Newly-Weds	131.9	131.1
	Newly Divorcees	45.1	30.9
Step 3	Indirect Identifiers	676.9	352.8

One particular group of taxpayers that proved problematic in linkage are people with irregular employment histories (such as those engaged in study), who do not submit a tax return every year. For this reason, it is not sufficient

to link only consecutive years; rather non-consecutive years must also be linked – for example, the years 2001 and 2003. From this, it was possible to identify a total of 313,000 taxpayers that submitted tax returns only in 2001 and 2003, and another 277,000 taxpayers who submitted a tax return in 2004 but not in 2003.

2.2 Marriage and Divorce

Marriage and divorce lead to at least one of the involved parties changing their tax number, it is necessary to identify some of the incoming and departing individuals that satisfy this case.

2.2.1 Marriage

Upon marriage, the wife takes on the tax number of her husband and both are together assessed as a so-called splitting-case. To identify these marriage cases, the cases which changed from being assessed as single in one year (t) and assessed as a splitting-case in the following year ($t + 1$) were first identified. These cases were then matched with the singles that could not be found in the $t + 1$ data. As matching identifiers the place of residence, religion and the birth date were used. Therefore, it is assumed that the newly-weds did not move or change their religion after marriage, an assumption that might not hold in every case.

Using this procedure about 132,000 taxpayers could be identified that married in 2001 and therefore did not hand in a personal tax claim in 2002. For the following years this number was 131,000.

This step of the algorithm is yet to be improved further. In particular, there are some cases where the newly-wed received a new tax number after marriage. It is also possible to match these, via indirect identifiers, but this has not been done at this point.

2.2.2 Divorce

As opposed to the marriage cases, after divorce one taxpayer becomes divided into two cases. Generally, he keeps the old tax number while she receives a new one. Similar to the procedure described in section 2.2.1 the so-far unmatched singles in year $t + 1$ become matched with the cases that were assessed as a splitting-case in t and as a single case in $t + 1$ by indirect identifiers.

The indirect identifiers used were the same as in 2.2.1 where the assumption that neither of them moved might be even less true than after a marriage.

Using this procedure about 45,000 divorcees could be identified in 2002 and about 31,000 in 2003.

2.3 Indirect Identifiers

At this point, after linking by using the tax number and after identifying the newly-weds and newly-divorcees, it is certain that not all of the unlinked taxpayers are part of the panel mortality. Especially since the cases where the tax number changed and the old one was not kept on file have not been yet examined. Therefore, in a third step these taxpayers are linked by using indirect identifiers. Hereby, it is necessary to look at the married and unmarried taxpayers separately since linking the married couples is easier than the linking the singles as will be shown below.

2.3.1 Married couples

The identifiers used here are the birth dates of both partners and their religion(s). The probability that two married couples have exactly the same birth dates is already considerably low: especially if one considers that the majority of taxpayers were already linked in the first step. To be extra certain, religion was used as an additional identifier.

Using this, it was possible to link 130,000 taxpayers in 2001/2002 and 105,000 taxpayers in 2002/2003. About 50,000 of the 130,000 moved between 2001 and 2002 which is why they received a new tax number (between 2002 and 2003: 19,500 of the 105,000 taxpayers).

2.3.2 Singles

For singles the linking procedure is much more difficult, since the previously-used identifiers of birth date and religion are ambiguous. The probability of two people having the exact same birth date is somewhat higher than that of two couples having the same birth dates. Furthermore, singles are more likely to be part of the panel mortality than married couples. This also explains why there are more singles not linked than married couples in the matching algorithm up to this point.

To overcome this problem, more indirect identifiers need to be applied and two additional steps were carried out. In the first new step it is assumed that the taxpayer did not relocate, since there are also other reasons for a change in tax number. In the second step, relocating is considered.

Sample identifiers are:

- Place of residence
- Religion
- Gender

- Business ID Number
- Birth dates of the first three children
- Source of income
- Type of employment.

Using these about 560,000 taxpayers could be linked between 2001/2002 and 248,000 between 2002/2003.

3. The Taxpayer-Panel

Containing the years 2001, 2002 and the preliminary data of 2003, the panel contains about 27.8 million taxpayers, when entries and exits to the panel are included as well as the taxpayer with an unsteady income. About 20 million taxpayers could be examined in all three years. The majority could be linked by using the income tax number. Approximately, 1 million cases were linked by using indirect identifiers, while marriages and divorces accounted for about 340,000.

Examination of the incomes of the taxpayers in the panel shows that both the mean and the median are higher in the panel than in the original data. This means that the incomes of taxpayers included in the panel are higher than those incomes not included (Table 2). Considering that most of the remaining taxpayers that could not be linked were those with lower incomes and either entering or leaving the workforce, the results of the panel look promising.

Table 2

Descriptive Statistics of the Panel

		Number of taxpayers	Mean	Median	Variation coefficient	Bottom 99%	Top 99%
		in 1000	in Euro				
Yearly Income Tax Statistics	2001	27890	34609	26682	347	34725	34843
	2002	27628	34541	26912	335	34484	34598
	2003	26531	34502	26960	402	34433	34572
Panel Data	2001	25401	36642	27510	339	36579	36706
	2002	26883	35500	27219	332	35441	35558
	2003	23662	36371	28339	401	36294	36449

3.1 Non-linked Taxpayers

According to the matching procedure described in section 2, 2.4 million tax-payers remain that could not be matched with anyone in the years 2002 and 2003. From the year 2002 there were 435,000 taxpayers and for the year 2003 1.1 million.

The following section will examine whether or not this is the true panel mortality.

Figure 2 shows the age of those non-linked cases from the year 2002 di-vided into married couples and singles. The majority of singles that could not be linked were in their early twenties, which is approximately the age of most entrants into the workforce. This leads to the conclusion that they are part of the true panel mortality.
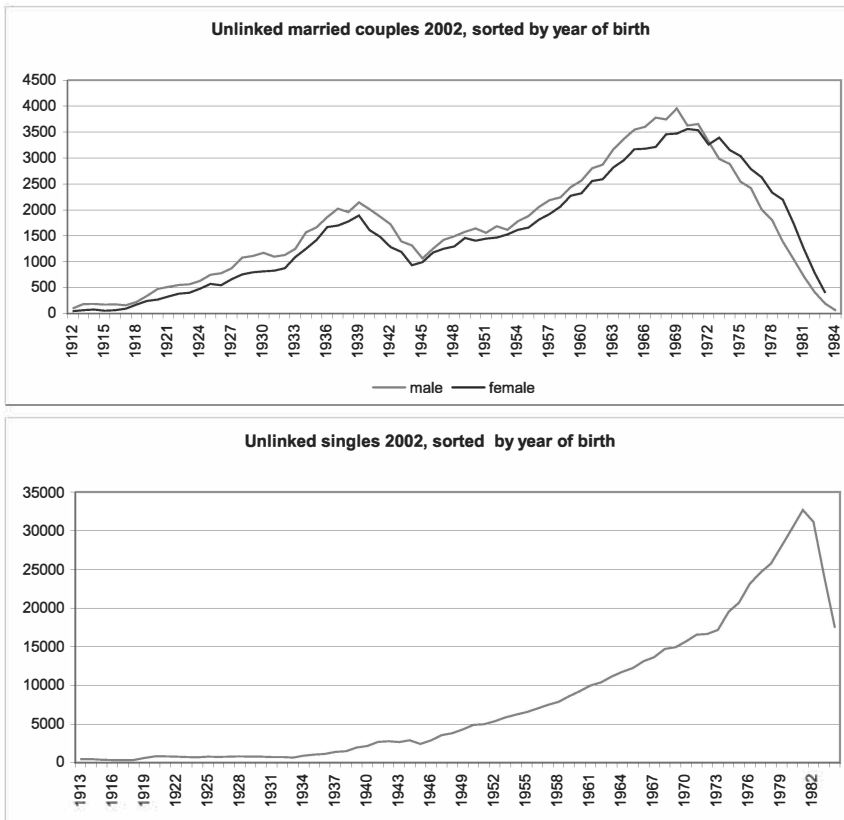

Figure 2

Conversely, the majority of married couples are between 30 and 35 years old. Considering that the average marriage age is 35.4 years (women 32.3 years), in 2002 the results are showing the newly-weds who handed in their first shared income tax claim. This also means that not all newly-weds could be identified and that more work is required.

The relative maximum at age $60-65$ was to be expected, since this is close to the retiring age.

4. Potential Analyses

In this section, the potential of the taxpayer-panel will be shown on some simple examples.

To maintain comparison, newly-weds, newly-divorcees and taxpayers with an unsteady employment history (the ones who did not hand in a tax claim for every year) were excluded. Additionally, only the years 2001 and 2002 are used; except for in section 4.1 where the analysis is based on all 3 years. Income in the following is defined as the adjusted gross income (sum of all incomes minus operating and income-related expensy). A taxation change on capital income was incorporated in the model[2].

4.1 Income Mobility

Income mobility is a typical field panel data can be used for. So far – in the absence of panel data – the income distribution was shown by dividing taxpayers into income classes. After comparing different years, there appeared to be little income variation. Whereas with panel data, it became possible to look at how the individual incomes change. Table 3 shows the volatility in income, where it can be seen that the high incomes are particularly volatile. As mentioned above, all three years were used. Excluded are taxpayers who had zero income in one particular year.

Since taxpayers at the upper or lower end of an income-class only need a small change in income to move into the class below or above, it is necessary to look at the actual change in individual income. Table 4 shows these income changes in percent. As can be seen, 12.2 million taxpayers (60 % of the panel) had an increase in income between 2001 and 2002. Between 2002 and 2003 this was true for about 11.6 million taxpayers (57 %). For about one third this increase was 5 % or less.

---

2 See Federal Statistical Office Germany (2006) for further details.

Table 3: **Taxpayers by Income**

Income range (EUR)	Panel Evaluation: 2001 – 2003						
	Population			Entries		Exits	
	2001	2002	2003	2001 – 2002	2002 – 2003	2001 – 2002	2002 – 2003
	in 1000						
losses < 0	218	238	243	236	211	89	159
0 – 10,000	2,649	2,700	2,893	864	972	961	836
10,000 – 20,000	3,428	3,427	3,379	1,337	1,264	1,338	1,311
20,000 – 30,000	4,457	4,437	4,317	1,436	1,324	1,457	1,443
30,000 – 50,000	5,670	5,695	5,653	1,351	1,259	1,326	1,301
50,000 – 75,000	2,698	2,722	2,712	753	704	729	714
75,000 – 100,000	830	851	864	341	325	319	312
100,000 – 150,000	438	444	453	182	177	176	167
150,000 – 200,000	121	121	123	70	69	70	66
200,000 – 250,000	53	52	52	35	34	36	34
250,000 – 375,000	51	49	48	28	26	30	27
375,000 – 500,000	18	17	16	12	11	13	11
500,000 – 1.000,000	20	17	16	9	8	12	9
1.000,000 – 2.500,000	8	6	6	3	3	5	4
2.500,000 +	3	2	2	1	1	2	1
Total	20,662	20,778	20,777	6,658	6,388	6,563	6,395

Table 4

Taxpayers by Change of Income

change from … to under	Years	
	2001 – 2002	2002 – 2003
	In 1000	
positive change		
0 % – 5 %	4 717.4	4 776.5
5 % – 10 %	2 159.5	2 071.0
10 % – 25 %	2 327.1	2 153.5
25 % – 50 %	1 204.8	1 084.5
50 % +	1 794.8	1 518.2
total	12 203.6	11 603.7
negative change		
0 % – 5 %	2 378.6	2 709.5
5 % – 10 %	1 214.1	1 271.7
10 % – 25 %	1 860.7	1 875.1
25 % – 50 %	1 392.0	1 410.3
50 % +	1 250.8	1 374.1
total	8 096.3	8 640.6

4.2 Progression and Horizontal Equity

The principle of horizontal equity, under which taxpayers with an identical income would pay the same amount in taxes, is a measurement to estimate whether or not the tax burden is equally distributed across the population. It is already questionable if this holds in just one year, but it certainly does not hold for several years. Taxpayers who can equally distribute their income over several years are in a better position than those who cannot.

For example, a married couple that earns €100,000 in two years could pay anything from €19,000 to €46,000 in tax. The difference of €27,000 is referred to as the progression effect. Examples such as this do not need empirical evidence, but to calculate the total progression-effect panel data is needed. To do this the actual income tax paid in the years 2001 and 2002 is compared with the fictional tax paid if those incomes could be equally spread over those two years. The results are shown in Table 5.

In 2001 the German government earned an extra €3,703 million in tax from people with large income changes. While the majority of taxpayers are only slightly effected by progression (as the median of about 4€ shows), there are some that suffer more. For about 1 % of German taxpayers the progression effect is €2,516 or more. It can be shown further that a smaller maximum tax rate lowers the progression effect. By running the analysis with the tax rates

from 2005, where the maximum tax rate was lowered from 48 % to 42 %, instead of 2001, the progression effect reduced.

Table 5

**Additional Tax Burden Based on Irregular Acquisition Income
for 2001 – 2002**

| Tariff | Number of taxpayers | Progression loss | | | | |
| | | mean | Median | Percentile | | Sum |
	in 1000			90	99	
2001	24,662	150	4	416	2,516	3,703
2005	24,662	136	3	389	2,299	3,365

To determine who these people are, the correlation coefficients for different income classes are displayed in Table 6. As can be seen, the correlation of the middle income class groups is high, meaning their income does not change very much. While people with high income have larger income changes, so do the people with very small or negative income. Since the tax burden of small income earners is lower, the group who is mainly suffering from the progression effect are people with high incomes.

Table 6

Spearman-Correlation Coefficient

Income from ... to under ... EUR	correlation coefficient 2001 / 2002
losses < 0	0.06
0 – 10,000	0.35
10 000 – 20 000	0.41
20 000 – 30 000	0.51
30 000 – 50 000	0.62
50 000 – 75 000	0.57
75 000 – 100 000	0.41
100 000 – 150 000	0.38
150 000 – 200 000	0.23
200 000 – 250 000	0.16
250 000 – 375 000	0.21
375 000 – 500 000	0.12
500 000 – 1 000 000	0.2
1 000 000 – 2 500 000	0.21
2 500 000 +	0.29

4.3 The Effect of Moving Home on a Person's Income

In general, the taxpayer-panel makes it possible to analyse how certain factors determine a person's income. The factor of moving home is used as an example here.

The panel data show that about 60 % of taxpayers had an increase in income between 2001 and 2002, with the remaining 40 % suffering a loss in income. Table 7 shows that an average taxpayer had to suffer an income loss from €479, while the income of taxpayers who moved increased by €544. It is remarkable that many singles gained from a move while married couples suffered an income loss. Reasons for that could be that married couples move due to other reasons, for example the birth of a child, rather than for a better job. This particular analysis will not be pursued further at this point.

Table 7

Change in Adjusted Gross Income, in Comparison to the Previous Year
Panel Evaluation for 2001–2002

Singles / married	Number of taxpayers in 1000	Change in Income	
		Mean	Median
		in €	
		Taxpayers in total	
Singles	11,740	287	587
married couples	12,923	−1,188	510
all	24,662	−479	561
		Taxpayers who moved	
Singles	715	1,088	1,060
married couples	222	−1,230	615
all	938	544	841

Further applications

Other analyses that can be done with the taxpayer-panel are the evaluation of tax reforms or the use of tax deduction. One application would be, for example, an examination of the so-called Riester-Rente pension (a privately funded pension scheme that is supported by the government through direct savings and tax deductions). Via the taxpayer-panel it is possible to analyse the acceptance of such products.

Years with tax reforms are particularly interesting, since such a reform can be seen as a natural experiment which allows observation of the individual

reactions. Gottfried and Schellhorn, for example, used the tax reform of 1990 in their work to estimate the effect that a change in tax rates has onto the taxable income.

5. Conclusion and Proceedings

With only three years of data the panel is only in its fledging stage, but its potential could be already demonstrated here. With every additional year the panel becomes more significant. An extra bonus would be that the panel mortality could be better identified in a panel with more years.

For the future it is planed to optimise the matching algorithm even further, especially steps 2 and 3 of the matching algorithm could be improved.

References

Federal Statistical Office Germany (2006): Jährliche Einkommensteuerstatistik auf Basis der Geschäftsstatistik der Finanzverwaltung 2002, Fachserie 14, Reihe 7.1.1, Wiesbaden.

Federal Ministry of Finance, An ABC of Taxes, Berlin 2002.

Goolsbee, A. (2000): What happens when you tax the rich? Evidence from Executive Compensation, Journal of Political Economy 108 (2), 352 – 378.

Gottfried, P. / *Schellhorn,* H. (2004): Empirical Evidence on the effects of Marginal Tax Rates on Income – The German Case, IAW-Diskussionspapier 15.

Hsiao, C. (2005): Why Panel Data?, Institute of Economic Policy Research, Working Paper 05.33, http: // www.usc.edu / iepr.

Kriete-Dodds, S. / *Vorgrimmler,* D. (2006): Das Taxpayer-Panel der Einkommensteuer-statistik, Wirtschaft und Statistik 1, 77 – 85.

Lechner, M. (2001): Eine Übersicht über gängige Modelle der Panelökonometrie und ihre kausale Interpretation, Universität St. Gallen, Department of Economics, Discussion paper No. 2001 – 18.

Lietmeyer, V. / *Gräb* C. / *Vorgrimmler,* D. (2005): Jährliche Einkommensteuerstatistik auf Basis der Geschäftsstatistik der Finanzverwaltung, Wirtschaft und Statistik 7, 671 – 681.