Chapter 6

Linking Federal Estate Tax Records

Jenny B. Wahl, St. Olaf College

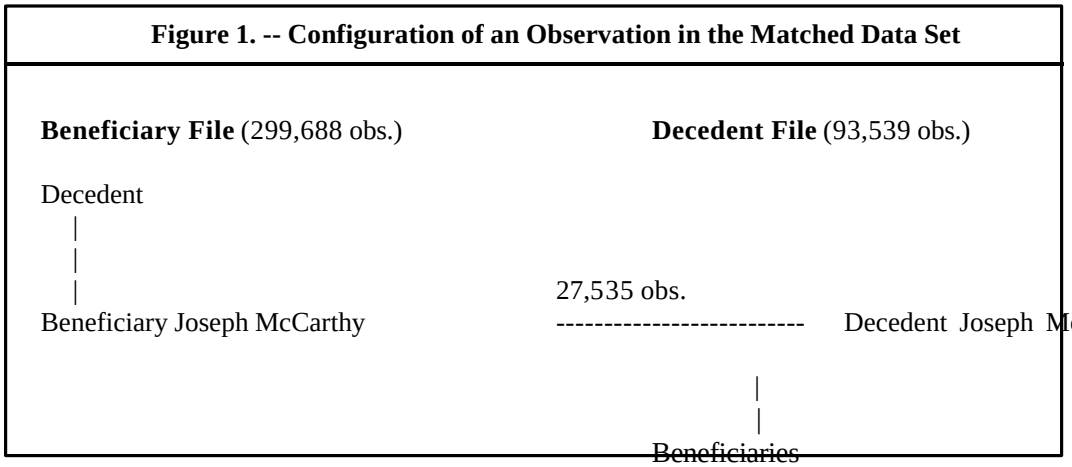
Abstract

This chapter focuses on the construction of a dataset that links together tax records and contemplates possible uses of these data. I first provide an overview of scholarly work regarding inherited wealth and establish the need for intergenerationally linked data. I then discuss techniques that I used to work with Federal estate tax returns filed in Wisconsin up to 1981 (which included 93,539 decedents and their 299,688 beneficiaries). By combining a standardizing/matching software package with a series of SAS programs, I linked these records to form a database containing 27,535 observations. Each observation has information on an individual who was reported on at least two estate tax returns: once as a decedent and at least once as a beneficiary. Of the 27,535 observations, 6,453 are matched pairs and the remaining 21,082 are likely pairs. I conclude by revealing certain problems associated with linking together tax records and by suggesting future research.

Introduction

The only sure things in life are death and taxes – and, unfortunately for some, death taxes. Fortunately for the rest of us, Federal estate tax data offer a rare opportunity to observe the total wealth, portfolios, and bequest behavior of certain individuals. Not only that, these data can be linked across generations, providing testing grounds for hypotheses about motives for intergenerational transfers, tradeoffs of family size and bequest amount, and the like. I have used all the estate tax records filed in Wisconsin from 1916 to 1981 to assemble just such a data set. These data consist of 27,535 observations; each observation has information on a single individual who was reported on two estate tax records: once as a decedent and once as a beneficiary [1]. Of this number, 6,453 are matched pairs and the remaining 21,082 are likely pairs. Figure 1 illustrates the configuration of each observation. In addition to the linked data, a residual set of 272,153 beneficiaries did not match to any decedent.

Figure 1. -- Configuration of an Observation in the Matched Data Set



What follows is, first, a brief overview of some of the questions and theories that scholars have put forth regarding wealth and intergenerational transfers. I then turn to a fuller description of the data, a discussion of the linking methodology, and a short mention of the empirical work that lies ahead.

The Importance of Inherited Wealth

For a variety of reasons, scholars have studied the transfer of wealth across generations. Some have focused on macroeconomic issues such as the influence of wealth transfers on the distribution of wealth (Menchik, 1979; Kotlikoff and Summers, 1981; Modigliani, 1988; and Tachibanaki, 1994), the degree to which intergenerational wealth transfers affect savings rates across countries (Darby, 1979; Hayashi, 1986; and Kotlikoff, 1988), and the interaction of cross-generational transfers and fiscal policy (Barro, 1974; and Aaron and Munnell, 1992). Others have concentrated on microeconomic questions such as the propensity of parents to compensate their less able children or, alternatively, to leave more money to their relatively capable offspring (Becker and Tomes, 1979; and Tomes, 1981).

In the process, researchers have speculated as to the appropriate model of behavior. Do individuals leave bequests because they care about their descendants or other heirs? Or do people design bequests strategically to induce potential heirs to offer attention and companionship? Or might the leaving of an estate simply be a mistake born of miscalculating one's own mortality? (Kotlikoff and Spivak, 1981; Bernheim, Schleifer, and Summers, 1985; Abel, 1985; Hurd, 1987; Modigliani, 1988; Lord and Rangazas, 1991; Altonji, Hayashi, and Kotlikoff, 1992; Gale and Scholz, 1994; Abel and Kotlikoff, 1994; Hurd, 1994; Arrondell, Perelman, and Pestieau, 1994; Yagi and Maki, 1994; and Tachibanaki and Takata, 1994.) Professors Martin David and Paul Menchik (1982) took yet a different tack. They used wealth data to estimate propensities to bequeath out of earnings. Although they did not propose any new theories, David and Menchik cast doubt on an old one: their results indicated that the life-cycle hypothesis cannot explain the bequest behavior of a set of Wisconsin decedents [2].

Others have posed additional interesting questions. Do people behave differently – choose alternative occupations or retire early, for example – if they receive or anticipate a bequest? What relationship do estate size and life insurance bear to a decedent's earnings? What connections exist among fertility, estate size, and earnings? Can one find evidence, for instance, of a tradeoff between the number of children and the wealth left to each one (Becker and Tomes, 1976; Behrman, Pollak, and Taubman, 1982; Wahl, 1986; and Wahl, 1991)? Do people tend to allocate estates equally among their children? Are people increasingly "spending the kids' inheritance," as the bumper stickers proclaim? What patterns in charitable giving have appeared over the years? Is age at death related to lifetime earnings? Many of these questions remain open. Answering them requires a sufficiently large, intergenerationally linked data set that contains comprehensive demographic and socioeconomic information.

The Original Estate Tax Data: Saved in the Nick of Time

Estate tax records contain a wealth of data on a nation's citizens. One can find not only detailed information on accumulated capital and portfolio holdings but also clues about family composition, residence and migration patterns, fertility, and mortality. By dint of much effort (and good computer software) one can even link records together to reconstitute families and their financial and demographic histories. I have drawn upon Federal estate tax records to do just this.

Let me offer a short history of the initial data collection effort. In 1916, the modern Federal estate tax came into being – well before modern computers, but recently enough that paper documents still existed in archives seven decades later. In efforts to clean house during the Reagan years, zealous politicians nearly caused an untimely end for the boxed estate tax returns that were scattered in warehouses around the coun-

try. Fortunately, the Statistics of Income (SOI) Division at the Internal Revenue Service marshaled its forces to preserve these important historical artifacts in computerized form. The result was two enormous files: one consisting of economic and demographic information on decedents, the other of information on beneficiaries (linked via record number to the original estate tax record).

Any attempt to match these two files required reducing their size. Because other researchers have used Wisconsin data to investigate wealth and estate issues (for example, David and Menchik, 1982), SOI extracted all the Wisconsin estate tax returns to use for a pilot project. The result was a decedent file with 93,711 observations and a beneficiary file with 300,269 observations. In the decedent file, 93,539 are unique individuals. For consistency's sake, omitting records from the decedent file meant purging the same records from the beneficiary file. The outcome was a file of 299,688 beneficiaries. Of this number, 188 seem to be duplicates on the same estate tax record – that is, beneficiaries with the same name and same relationship code to the decedent, but appearing twice on a given tax return. Such apparent duplicates may, however, represent different persons with the same name – cousins, for example. Alternatively, these may constitute separate bequests to a single individual – one direct and one in trust. Rather than investigate these observations before the match procedure, I simply marked them so that, if any appeared after the match, I could inspect them more carefully at that time.

Linking the Data: Overlapping Estate Tax Returns

Linking data from one set of records to another requires much information and, frequently, creative computer programming (Fellegi and Sunter, 1969). The AUTOMATCH software written by Matt Jaro provides a solid foundation (Jaro, 1997); variations on his programs coupled with SAS programming produced the linked estate tax records. The critical linkage was this: Joseph McCarthy, say, appears as a beneficiary on his father's estate tax return. In turn, the estate of Joseph McCarthy also files a tax return. The two are linked into a single observation, given consistency in social security numbers, sex, years of birth and death, and the like. Each observation then contains detailed information about the Joseph the decedent: his portfolio, age, marital status, and number of children, for instance. Information about Joseph the beneficiary appears as well: his relationship to his benefactor, receipt of a trust, and sometimes the size of his bequest.

The AUTOMATCH software contains several attractive features that help create good links between records. It standardizes individual names and creates NYSIIS and Soundex codes. (Because I had maiden names for many women, I ran the standardization/coding step twice.) These codes work well as blocking variables in the match process. The software also allows specification of values for missing variables; this helps distinguish between true mismatches and apparent mismatches caused by missing data. The match procedure itself allows multiple rounds so that I could block and match over different sets of variables. Table 1 shows the salient variables for each match round.

The matching process itself also has nice characteristics. I could request multiple matches -- important, because Joseph McCarthy may have inherited from more than one person. Each matching variable has a designation to control for miskeying in the original data. For example, I could allow for mismatched numbers in the social security number string and mismatched letters in the name character strings. These designations also allow matching around intervals, which proved essential for my year-of-birth variables because I had to construct them from rounded-year ages. Each matching variable also carries a set of probabilities to allow for type I and type II errors [3]. All together, these probabilities translate into a single weight associated with each match in each match round. I could choose two cutoff weights per round: one the lower bound for declared matches, the other the lower bound for potential matches. After each match round, I could perform an interactive clerical review on the potential matches and change their status to declared matches or residuals. Following the clerical review, the software outputs all residuals to the next match round.

Table 1. -- Matching Rounds

Match Pass	Blocking Variables	Matching Variables	Original Matches	Original Clerical	Final Matches	Final Clerical
1	SSN	surname first name maiden name suffix initial sex year of birth	4,805	119	4,884	0
2	SSN	surname/ maiden name first name suffix initial sex year of birth	4,906	43	4,928	0
3	surname NYSIIS first name Soundex sex	SSN surname first name maiden name initial suffix year of birth	5,514	30,651	5,514	30,651
4	surname/ maiden name NYSIIS first name Soundex sex	SSN surname/ maiden name first name initial suffix year of birth	5,515	30,652	5,515	30,652

The clerical review process is extremely time-consuming. Although I used it for the first two match rounds, thereafter I used SAS programs to decide whether to change the status of potentially matched pairs [4]. Simply put, I distilled a set of decision rules into SAS programs rather than using the same rules on an interactive, case-by-case basis. For example, suppose the initial matching process paired Joseph McCarthy from the decedent file to Joseph McCarthy from the beneficiary file. The beneficiary file includes a date of death for the Joe's benefactor. If this date of death was after the date of death of Joe the decedent, I called it a nonmatch.

Particular Features of Estate Tax Data

Any two data sets have quirks that make matching difficult. Let me point out a few issues associated with matching data on people observed at two different points in time, often several years apart.

Some problems pertained primarily to females. During the time period covered by my data, a woman often took her husband's social security number at marriage. Sorting and matching by SSN for women was therefore problematic if a woman got married after receiving a bequest. Women also sometimes changed their middle initials upon marriage to reflect their maiden names. I had to take care, then, with the probabilities placed on type I and type II errors when initials appeared as matching variables.

Yet women provided information – namely, maiden names -- that helped me refine the likelihood of matches as well. Suppose a decedent carried the maiden name Scheuren. Say that the decedent potentially matches to a beneficiary, whose benefactor carried the last name Scheuren. Provided that birth and death years were logical, I could declare this a match. By the same token, if a (potentially matched) decedent had the last name Winkler and the benefactor named on the beneficiary file had the maiden name Winkler, again this might be considered a match.

Males created certain problems as well, albeit less directly than females. I had hoped to use cities as matching variables. Yet this hope was dashed: Wisconsin men seemed to like passing their names on to their sons, people did not seem to move around much, and missing ages for beneficiaries frequently meant that I could not screen matches by birth year. As a result, I could not use locational variables to improve the matching process.

A last discovery: one should always assign unique record identification numbers to observations on each file. Initially, the beneficiary file contained identifiers that pointed back to the estate tax record, but it did not have unique identifiers. Because my original files were so large, I excluded some variables while performing the match. When I attempted to reattach data after the match, I could not be sure that the right data went to the right individual. I therefore had to retrace my steps, this time with unique identifiers for each original file.

What Lies Ahead

In the coming months, I will use these linked data to fulfill two objectives. One is to compare matched and unmatched beneficiaries and report any significant differences. The other is to generate a proxy for bequest amount. To proceed, I must convert dollar figures to constant-dollar amounts, control for changes in filing thresholds, and implement a logical cutoff process so as to separate nonmatches from impossibilities. That is, I do not want to call unmatched data a "nonmatch" if the individual could not possibly have entered the matched data set because he or she was born before 1916 or was still living after 1981. Eventually, I hope to extend matches forward and back to reconstitute multiple generations of families.

Acknowledgments

Many thanks to Fritz Scheuren, who made sure this work saw the light of day; Dan Skelly, who encouraged me to dirty my hands with these data; and Barry Johnson, who answered my questions and did most of the hard stuff.

Footnotes

[1] Individuals can appear as beneficiaries on more than one estate tax return. The pairs do not therefore represent unique persons.

[2] The life-cycle hypothesis, associated originally with Franco Modigliani, suggests that people tend to decumulate wealth after a certain age, as they begin to anticipate death. For a review, see Ando and Modigliani (1963) and Modigliani (1988).

[3] Type I errors occur when true matches are declared nonmatches; Type II errors occur when nonmatches are declared matches.

[4] Here is a time comparison: using the clerical review process on 3,827 potential pairs took me seven hours. Writing and running SAS programs with embedded decision rules took about one-half hour for the same data.

References

Aaron, H. and Munnell, A. (1992). Reassessing the Role for Wealth Transfer Taxes, *National Tax Journal,* 45: 119-44.

Abel, A. (1985). Precautionary Saving and Accidental Bequests, *American Economic Review,*75: 777-91.

Abel, A. and Kotlikoff, L. (1994). Intergenerational Altruism and the Effectiveness of Fiscal Policy – New Tests Based on Cohort Data, *Savings and Bequests*, ed. T. Tachibanaki, Ann Arbor: University of Michigan Press, 167-96.

Altonji, J.; Hayashi, F.; and Kotlikoff, L. (1992). Is the Extended Family Altruistically Linked? New Tests Based on Micro Data, *American Economic Review,* 82: 1177-98.

Ando, A. and Modigliani , F. (1963). Lifecycle Hypothesis of Savings: Aggregate Implications and Tests, *American Economic Review,*53.

Arrondell, L.; Perelman, S.; and Pestieau, P. (1994). The Effect of Bequest Motives on the Composition and Distribution of Assets in France, *Savings and Bequests*, ed. T. Tachibanaki, Ann Arbor: University of Michigan Press, 229-44.

Barro, R. (1974). Are Government Bonds Net Wealth? *Journal of Political Economy,* 82: 1095-1118.

Becker, G. and Tomes, N. (1976). Child Endowments and the Quantity and Quality of Children, *Journal of Political Economy,* 84: 143-62.

Becker, G. and Tomes, N. (1979). An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility, *Journal of Political Economy,* 87: 1153-89.

Behrman, J.; Pollak, R.; and Taubman, P. (1982). Parental Preferences and Provision for Progeny, *Journal of Political Economy,* 90: 52-73.

Bernheim, B. D.; Schleifer, A.; and Summers, L. (1985). The Strategic Bequest Motive, *Journal of Political Economy,* 93: 1045-76.

Darby, M. (1979). The Effects of Social Security on Income and the Capital Stock, Washington, DC: American Enterprise Institute.

David, M. and Menchik, P. (1982). Modeling Household Bequests, University of Wisconsin, working paper.

Fellegi, I. and Sunter, A. (1969). A Theory for Record Linkage, *Journal of the American Statistical Association,* 64: 1183-1210.

Gale, W. and Scholz, J. K. (1994). Intergenerational Transfers and the Accumulation of Wealth, *Journal of Economic Perspectives,* 8: 145-60.

Hayashi, F. (1986). Why is Japan's Saving Rate So Apparently High, *NBER Macro Annual,* ed. S. Fisher, Cambridge: MIT Press.

Hurd, M. (1987). Savings of the Elderly and Desired Bequests, *American Economic Review,* 77: 298-312.

Hurd, M. (1994). Measuring the Bequest Motive: The Effect of Children on Saving by the Elderly in the U.S., *Savings and Bequests*, ed. T. Tachibanaki, Ann Arbor: University of Michigan Press, 111-36.

Jaro, M. (1997). Matchware Product Overview, *Record Linkage Techniques – 1997*, eds. W. Alvey and B. Jamerson, Washington, D.C.: Office of Management and Budget.

Kotlikoff, L. (1988). Intergenerational Transfers and Savings, *Journal of Economic Perspectives,* 2: 48-51.

Kotlikoff, L. and Spivak, A. (1981). The Family as an Incomplete Annuities Market, *Journal of Political Economy* , 89: 372-91.

Kotlikoff, L. and Summers, L. (1981). The Role of Intergenerational Transfers in Aggregate Capital Accumulation, *Journal of Political Economy,* 89: 706-32.

Lord, W. and Rangazas, P. (1991). Savings and Wealth in Models with Altruistic Bequests, *American Economic Review,* 81: 289-96.

Menchik, P. (1979). Intergenerational Transmission of Inequality: An Empirical Study of Wealth Mobility, *Economica* , 46: 749-62.

Modigliani, F. (1988). The Role of Intergenerational Transfers and Life Cycle Saving in the Accumulation of Wealth, *Journal of Economic Perspectives,* 2: 15-40.

Tachibanaki, T, ed. (1994). *Savings and Bequests*, Ann Arbor: University of Michigan Press.

Tachibanaki, T. and Takata, S. (1994). Bequest and Asset Distribution: Human Capital Investment and

Intergenerational Wealth Transfers, *Savings and Bequests*, ed. T. Tachibanaki, Ann Arbor: University of Michigan Press, 197-228.

Tomes, N. (1981). The Family, Inheritance, and Intergenerational Transmission of Inequality, *Journal of Political Economy*, 89: 928-58.

Wahl, J. (1991). American Fertility Decline in the Nineteenth Century: Tradeoff of Quantity and Quality? *Essays in Honor of Robert William Fogel*, eds. C. Goldin and H. Rockoff, Chicago: University of Chicago Press.

Wahl, J. (1986). New Results on the Decline in Household Fertility in the United States from 1750 to 1900, *Studies in Income and Wealth*, eds. R. Gallman and S. Engerman, Chicago: University of Chicago Press, 391-438.

Yagi, T. and Maki, H. (1994). Cost of Care and Bequests, *Savings and Bequests*, ed. T. Tachibanaki, Ann Arbor: University of Michigan Press, 39-62.