Wealth mobility and dynamics over entire individual working life cycles*

Stefan Hochguertel Henry Ohlsson

February 1, 2010

PRELIMINARY AND INCOMPLETE

Abstract

We study taxable wealth in unique Swedish administrative data, annually following a large sample of households over a period of almost 40 years. The main data limitation is non-observability of wealth for those below the tax exemption level. We exploit the long panel dimension by estimating dynamic 'fixed effects' models for limited dependent variables that allow for individual heterogeneity in both constants and autoregressive parameters, and control for heterogeneity through observables. We find substantial wealth mobility over the long time spans, partly accounted for by life-cycle behavior, while sufficiently capturing dynamics by an AR(1) process at the individual level.

Keywords: wealth mobility, wealth dynamics, life cycle, heterogeneity, panel data

EconLit subject descriptors: C230, D140, D310, D910, H240

Stefan Hochguertel, Department of Economics, VU University Amsterdam and Tinbergen Institute, The Netherlands, <shochguertel@feweb.vu.nl>

Henry Ohlsson, Department of Economics, Uppsala University and Uppsala Center for Fiscal Studies (UCFS), Department of Economics, Uppsala University, Sweden, <henry.ohlsson@nek.uu.se>

*We are grateful for financial support from the Nordic Tax Research Council. Helpful comments and suggestions from Martin Browning, participants at the Danish Microeconometric Network Meeting, July 2009, Copenhagen, and seminar participants at VU University, Amsterdam and Uppsala University are gratefully acknowledged. Some of the work was done when Ohlsson enjoyed the hospitality of LEM, Université Panthéon-Assas, Paris II.

1 Introduction

Who becomes wealthy? Who stays wealthy? And who will always remain poor? The opportunities to accumulate wealth create incentives for education, work effort, and entrepreneurship. We would expect considerable wealth mobility if these incentives are strong and affect behavior. As people differ in many respects, we would also expect to see considerable heterogeneity in wealth trajectories.

We study movements of individuals and households in the wealth distribution over time and, therefore, as they age in this paper. The data available allow us to track households' wealth transitions over most of their working lives. This makes our data unique. Those getting rich not only increase their wealth over time in an absolute sense, but they also move through the wealth distribution and improve their position in the wealth ranking. Wealth distributions are highly skewed. For instance, the top percent of households owns about one third of private net worth in the US. This fact makes it necessary to capture the top percentiles in a reliable way. It is a strength that the data we use meet this requirement.

The degree of intragenerational wealth mobility is important when discussing different economic issues. First, wealth accumulation is the result of *choices* concerning labor supply, consumption, and savings. Life-cycle models predict that individuals will accumulate wealth while working and then decumulate when retired. One set of issues concern how well the life-cycle model predicts the actual age-wealth profiles and if these profiles differ between individuals. Another issue is if controlling for other determinants of wealth reduces the observed heterogeneity in age-wealth profiles. While we can, in principle, control for education, other important determinants of wealth accumulation such as entrepreneurial ability, are inherently unobservable.

Second, wealth mobility reflects the extent to which there is *equality of opportunity* in a society. If there is equality of opportunity, the wealth of a young person will not be a good predictor of this person's wealth when middle aged. Suppose that entrepreneurship and risk taking sometimes for some yield considerable wealth increases. If wealth taxation reduces entrepreneurship and risk taking, we would then expect reduced wealth mobility. Wealth during different phases of the life cycle will be highly correlated if, on the other hand, inherited wealth is important. Inheritances are, however, very unequally distributed. This means that if inheritances are important for wealth then inheritances will be a source of heterogeneity.

Finally, from a macroeconomic perspective, more and more modeling efforts are being spent on accommodating agent *heterogeneity* into models explaining consumption and saving behavior, investment, or business cycles. Recent literature tries to escape from the straightjacket assumptions and implications of representative agent economies.[1] The degree of heterogeneity in dynamic wealth accumulation appears to be unknown, however, judging from macroeconomic studies

[1] See, for instance, Browning et al. (1999) and Bertola (2000).

that calibrate models to moments of the cross sectional wealth distribution.[2] We provide descriptive evidence on heterogeneity in dynamic wealth accumulation, evidence that can be compared to the properties of calibrated models.

The previous literature on wealth mobility includes Hurst et al. (1998), Jianakoplos and Menchik (1997), Keister (2005), and Steckel and Krishnan (2006) who all study wealth mobility in the US. Jappelli and Pistaferri (2000) study wealth mobility in Italy. Klevmarken et al. (2003) and Klevmarken (2004) are among the previous papers on wealth mobility in Sweden.

These studies are based in wealth observations, in the time dimension, for 2–4 years. Wealth mobility is studied by comparing individual households' positions in the wealth distribution, in most cases, 5–7 years apart. Sometimes the time span is down to 2 years, sometimes up to 10–15 years apart. The sample sizes are quite small, in the cross-section dimension there are observations for 1,000–5,000 households. Wealth mobility is often coarsely defined as movements between quartiles, quintiles, or deciles in the wealth distribution.

Most studies find that the probabilities to stay poor and remain rich are comparatively high. Wealth mobility is predominantly high in the middle of the wealth distribution. The previous literature consists of single country studies. Klevmarken et al. (2003) is the only exception. This paper compares wealth mobility in the US and Sweden. Contrary to what many might have conjectured, Klevmarken et al. (2003) find that wealth mobility in Sweden is as high as in the US.

The previous literature is, however, limited by the small number of observations. In the time dimension, the few observations for specific individuals for different years can only account for very limited parts of the individual's life cycle. In the cross section dimension, the few observations of different individuals for a specific year means that observations can only be grouped into a few quantiles. This implies that the measure of mobility becomes imprecise when mobility is defined as movements between quantiles. With few individuals it also becomes difficult to study patterns in individual heterogeneity. These limitations also reduce the possible choices of empirical methods to study mobility. In addition, the previous literature is based on survey data. Surveys tend not to do so well in covering the top percents of the wealth distribution.

We believe that we can deal with these shortcomings of the previous literature. The data available to us are from the LINDA data base, an administrative source from Statistics Sweden. This data base provides long individual time series, many individuals, and the top percents of the wealth distribution well documented. This enables us to improve considerably on the analysis of wealth mobility.

The LINDA data base includes 3 percent of the Swedish population and their household members. There are 300,000 households and 700,000 individuals in this data base. We can follow a considerable part of individual life cycles for many. There are close to 40 annual observations for some individuals.

The key variables we use are annual taxable net wealth at the individual level

[2]Examples are Huggett (1996), Castañeda et al. (2003), and Cagetti and De Nardi (2009).

and at the household level from 1968 and onwards.[3] A main advantage with this data set is that for those who do pay taxes there are very precise wealth measurements available.[4]

This means that our measure of wealth mobility is very closely related to whether or not the individual pays wealth taxes. Wealth mobility is interpreted as the movements in and out of the top percents of the wealth distribution over time and, also, movements over time within the top percents. As an alternative we also use an absolute real wealth measure, movements across a real wealth threshold.

The very long individual time series allow us to study "individual wealth trajectories", at least for those who pay wealth taxes at some stage. Accounting for such detail on individual heterogeneity has, to our knowledge, not been done and not been possible before in this context.

We present empirical estimates of nonlinear dynamic panel data models controlling for measurable determinants of wealth. At the same time we allow for individual (household) heterogeneity in both constants and autoregressive coefficients that describe short-run dynamics. Our main results are:

- We find results indicating a very large degree of heterogeneity in wealth trajectories. The autoregressive coefficients vary substantially in the population. There is, in other words, considerable heterogeneity even among those in the top percentages in the wealth distribution.
- We also find considerable movements into and within the top percents in the wealth distribution. This is not quite consistent with previous results for Sweden presented by Klevmarken (2004).

The rest of the paper is structured as follows: Section 2 presents our theoretical framework. In Section 3, we present the data and how the data set was constructed. We also present some descriptive results in this section. Section 4 presents our econometric approach. The evidence from the main specifications can be found in Section 5. Section 6 concludes.

2 Theoretical framework

The objective of this section is to provide a theoretical framework for studying wealth and wealth accumulation. We will discuss the various determinants and sources of wealth (or its absence).

2.1 Determinants of wealth accumulation and wealth heterogeneity

Think of a young adult in her early or mid 20's. When starting out in working life she has been given some *initial conditions* provided by her parents. There are

[3] Taxable wealth at the household level was also the actual tax base during the studied period.

[4] A disadvantage is that wealth information in the register data is only available for those whose taxable wealth exceeds the high tax exemption levels.

four main ways by which parents can make transfers to their children: First, there are biological transfers of natural talents and abilities (genes). Second, parents can also transfer financial and tangible property by *inter vivos* gifts and bequests. For our young adult these intergenerational transfers are probably expected rather than already realized. Third, parents can contribute to the formal education and other human capital investments of the child. Finally, parents can provide 'social capital', for example, values, manners, and access to social networks.

Parents are different and transfers will differ. The transfers from parents will, therefore, create an *initial heterogeneity* among young adults entering working life. Family background will, in other words, be important for, among other things, wealth and wealth accumulation. We are here talking about conditions like parents' education, occupation, and marital status. Family size and family income and wealth are also important family background characteristics. Culture, religion, race, and ethnicity are also characteristics that have been mentioned in the literature.

Gender and country of birth are other characteristics that contribute to initial heterogeneity. It may also be important to which birth cohort the individual belongs. Birth cohorts differ in size, but things like the date of labor market entry may also differ between cohorts for exogenous reasons.

Given the initial conditions our young adult will make *choices* and continue to do so during her life. Her preferences–for example, her time preference rate and her risk attitude–will be important for her choices. One of the outcomes will have to do with the path of her working life. Important dimensions of this are hours of work, occupation, career path, and entrepreneurship.

Another decision is the consumption path over the life cycle. The optimal consumption path will not necessarily follow the income path. Life cycle saving in general and retirement saving in particular will follow from the choices made. The future savings of our young adult might also be affected if she wishes to leave a bequest or if she, because of uncertainty, saves for precautionary reasons.

This will, of course, result in wealth accumulation and decumulation over the life cycle. But wealth might also be affected by the investment behavior of the individual, for example, the portfolio composition. (The term financial literacy has been used in the literature.)

The time and age pattern of demographic choices will also affect wealth. Marital status, family size, and the number of children are important characteristics.

Our young adult might be lucky or unlucky during the course of life. Windfalls such as unexpected inheritances, lottery winnings, and gambling winnings will increase wealth, at least temporary.

But windfalls might affect many and not only specific individuals. Asset prices might move so that the wealth of many is affected simultaneously. This is one example of how *general economic conditions* might affect wealth. The taxation of wealth is another example. The differences between living in different geographical locations may also change over time.

With this sketch of the factors that might affect wealth and wealth accumulation, we will now turn to a more formal discussion of the individual's life cycle choices.

2.2 Optimal wealth accumulation

The objective of this subsection is to discuss the implications for wealth accumulation of the choices the individual makes concerning consumption and savings.[5] The approach is to start by focusing on the modeling assumptions needed to have individuals making the same choices rather than different choices.

Suppose that there is no uncertainty. Individuals have the same length of life and no bequest motives. They meet the same constant rate of interest. Each household consists of a single individual. Utility is additively separable, the instantaneous utility function does not change over time, and the time preference is constant.

The individuals maximizes

$$U = \sum_{t=1}^{T^*} \frac{u(C_t)}{(1+\rho)^{t-1}}, \tag{1}$$

where U is utility, u is instantaneous utility with decreasing marginal utility, t is time, T^* is the length of life, C is consumption, and ρ is the time preference, by choosing a consumption path $C_t,\ t = 1,\ldots,T^*$ subject to the intertemporal budget constraint

$$\sum_{t=1}^{T^*} \frac{C_t}{(1+r)^{t-1}} = \sum_{t=1}^{R} \frac{E_t}{(1+r)^{t-1}} + W_0, \tag{2}$$

where r is the rate of interest, R is the retirement age, E is earnings, and W_0 is the value of initial wealth in the beginning of period 1. The left hand side is lifetime consumption C^L, the right hand side lifetime resource consisting of lifetime earnings E^L and initial wealth. Provided that $R < T^*$, there will be retirement saving so that the individual can consume as retired. Consumption will be smoothed over the life cycle.

Let us add the following assumptions: Suppose that the interest and time preference rates are zero, that initial wealth is zero, and that annual earnings are constant during the individual's working life. The individual will choose to consume a fixed share of lifetime earnings every year. This will result in piecewise linear age-wealth profile with increasing wealth until retirement, a wealth peak at retirement, and then decreasing wealth. The wealth of individual i will evolve according to

$$W_{it} = W_{it-1} + (1 - D_i^R)\left(\frac{1}{R_i} - \frac{1}{T_i^*}\right)E_i^L - D_i^R \frac{1}{T_i^*} E_i^L, \tag{3}$$

where W is wealth and D_i^R is an indicator equal to one when individual i is retired and zero otherwise. The savings rate of a working individual is

$$s_{it} \equiv \frac{W_{it} - W_{it-1}}{E_i^L} = \frac{1}{R_i} - \frac{1}{T_i^*}. \tag{4}$$

[5] The discussion is inspired by Davies and Shorrocks (1999) and Dynan et al. (2004).

Suppose that individuals are identical except for age. During their working life individuals will move up in the wealth distribution both in absolute and relative sense, as retired individuals will move down.

It is an old question in the economics literature whether rich people save more than poor people. Dynan et al. (2004) discuss under which conditions savings rates are the same. Savings rates provide a link between income and wealth. Suppose that individuals have different lifetime earnings while there is no uncertainty and there are no bequest motives. With identical savings rates for a cohort j, the wealth of an individual belonging to the cohort will evolve according to

$$W_{ijt} = W_{ijt-1} + s_{jt} E^{L}_{ij}. \tag{5}$$

The cohort specific savings rate is s_{jt}. Consumption is proportional to lifetime earnings for the individual either if (i) the time preference rate is constant and equals the rate of interest or if (ii) preferences are homothetic. In the first case annual consumption will be same every year, in the second case annual consumption will grow at the same rate every year. In addition, suppose that preferences, length of life, and rates of interest are the same for all individuals. The ratio of consumption to lifetime earnings at time t is the same for all individuals belonging to cohort j. Finally, suppose that the relative differences between individuals in annual earnings are constant over time. The savings rate at time t will then be the same for all individuals belonging to cohort j with these assumptions. There will, in other words, be no cross section variation at time t for those of the same age. The savings rate might, on the other hand, vary over time (age) for a given cohort. During their working life individuals will move up in the wealth distribution both in absolute and relative sense. Those with higher lifetime earnings will move faster and end up with more wealth at retirement than those with lower lifetime earnings.

Relaxing any of these assumptions and instead introducing, for example, differences in preferences or earnings profiles, rates of interest, length of life, retirement age, or introducing uncertainty and bequest motives will result in less homogeneity across individuals in wealth accumulation.

2.3 Framework for empirical specifications

Going from these simple theoretical models to model specifications for empirical analysis, keeping a life cycle perspective, allows for a host of possible choices that have been discussed in the literature. See our musings above in Subsection 2.1. What all models must have in common, however, is that an individual (a household) has to obey its lifetime budget constraint (the present value of consumption cannot exceed the present value of income receipts from all sources and initial wealth).

Consider the following simple equation of motion that describes wealth dynamics for household i between two periods $t-1$ and t and is implied by life cycle accounting:

$$W_{it} \equiv (1 + r_{it-1}) W_{it-1} - \Theta(X_{it-1}, \tau_{t-1}, W_{it-1}) + E_{it} + Tr^{pu}_{it} + Tr^{pr}_{it} - C_{it} \tag{6}$$

where the rate of interest r is possibly household specific, Θ is the tax liability, itself a function of tax code parameters such as an exemption level X and (marginal) tax rates τ. Further, let Tr^{pu} and Tr^{pr} denote public and private transfers.

Equation (6) is formulated for the case that there is a single asset available, but can be rewritten to allow for wealth composition and returns that are specific to portfolio items. The rate of interest r may then be interpreted as a price-index-type of average return. It may be household specific since portfolio compositions are choices that reflect, among others, household risk attitudes. Conditional on income components and consumption, the equation (6) is autoregressive in W.

This is about the only prediction that we can generate without being more specific in terms of modeling income processes and a utility function (and implied consumption demand). The equation is a basic accounting relation and does not generate by itself additional insights in terms of economic behavior coming from optimization, preferences, income paths, and various shocks. Clearly, income and consumption dynamics will determine wealth dynamics, and hence, not only unobservables such as risk attitudes, time preference rates and habits may have repercussions for individual wealth trajectories, but also age patterns and productivity shocks over time and generations in earnings.

These remarks may suffice at the time being to motivate empirical work on estimating dynamic equations like (6) while allowing for substantial heterogeneity, where possible not only entering through C (as a 'fixed effect' individual constant), but also through the coefficient on W_{t-1}.

3 Data and descriptives

3.1 Data source

Our data are from the Longitudinal INdividual DAta base (LINDA), a data source collected and maintained by Statistics Sweden.[6] The source data are various administrative data bases from government agencies that keep records on any (registered) inhabitant in the country. For instance, data from the tax authorities, the social security administration, and from local municipalities. We have spent considerable energy in trying to get at coherent definitions of variables from an array of different variables for different years in the source data.

The data come in two sub-samples, that we want to refer to as the 'P' sample (the panel sample) and the 'F' sample (the family sample). For the 'P' sample, the data were randomly drawn in 1994 with a sample size of 300,000 households, comprising almost 700,000 individuals. A household in the data set is a group of people treated as a taxable unit. For the vast majority of cases, this coincides with a residential household or a family. All members of these 1994 households were then followed through time, backwards until 1968, and assigned the same household

[6] Edin and Fredriksson (2000) presents the data base.

number as the 1994 one if they were members of that same tax household in the respective year.

For those members who joined the 1994 households in other years, a different household number was assigned before joining. The data also tracks those 'joining' members through time when they are not member of a 1994 household. Likewise, the data were extended beyond 1994 until 1999, using a similar sampling scheme. This implies that the change in the number of households and individuals is closely following the development of the entire residential population in the country for the period 1968 through 1999.

The 'F' sample is available to us from 1991 until 2005. The sampling unit here is a 'family' that is, persons living at the same address. Since there may be various sub-households within a 'family' that are treated as separate taxable units, and since members of the same tax households may live at different addresses, it may be that the definitions of 'households' in the 'P' sample and of 'families' in the 'F' sample do not coincide. On average, a 'family' is slightly larger than a 'household'.[7]

The administrative nature of the data implies that there is no panel data attrition as is known from survey data. Theoretically, a person can leave the sample by emigration or death (and only in a few cases where records could not be traced in the source data bases). Persons enter by birth or by, say, marrying into an existing unit.

Following individuals over a time span of nearly 40 years inevitably implies that they live in different households of different composition at different stages of their life cycle. For instance, an individual might be born in household number 1, then complete school and start working and be separately taxable, so be assigned to household number 2, then marry, have children on their own, and subsequently divorce, upon which again a new household number 3 is assigned. The implication is that there are many 'households' that are linked on an individual level since the same person is in household 1 in one year and in household 2 or 3 in another year.

We aim to remove split-off households. For this, we first create a new super-household identifier that groups all individuals that ever were in a household that shared at least one member in anyone year. Within such a super-household, we select that household that ranks highest in average size and participation within the 'P' sample. We call this the 'core' household.

For the 'F' sample we create artificial units from the recorded families by tracing the existing and joining members of a 1999 core household that share a family

[7]Table A.1 in Appendix A fills in on the relative differences. It shows that the number of households virtually equals the number of families in any year of overlap, but that, on average, families are about 15 percent larger than tax households. Since in two thirds of all cases the same individuals form both a household and a family in any given year, and close to 99 percent of all individuals that are in the household data are also in the family data, we aim to combine both data sets and work, in what follows, with the smaller definition of 'tax households'. One large difference between the series occurs at the point in time when children of age 18 and above earn their own incomes and own their own wealth and are thus separately from their parental household liable to tax.

Table 1: Percentage of households paying wealth tax, 1968–2005

year	tax payer	year	tax payer	year	tax payer	year	tax payer
1968	6.22	1978	5.57	1988	11.91	1998	10.56
1969	6.59	1979	5.94	1989	13.09	1999	11.80
1970	4.12	1980	6.66	1990	5.44	2000	11.23
1971	5.13	1981	4.98	1991	5.96	2001	7.58
1972	5.58	1982	5.63	1992	6.94	2002	4.19
1973	5.95	1983	11.71	1993	8.13	2003	5.13
1974	3.64	1984	6.95	1994	6.69	2004	5.22
1975	5.79	1985	8.12	1995	7.28	2005	3.42
1976	6.15	1986	10.16	1996	7.94		
1977	6.48	1987	10.14	1997	9.82		
share ever paying wealth tax, 1968–2005							15.76
share ever paying wealth tax, those observed every year 1968–2005							34.05

Source: Linda, 1968–2005, full core sample.

identifier. We refer to these also as core households. Our subsequent analysis shall only consider core households.

The dependent variable we use is annual taxable net wealth at the household level. The tax base was a comprehensive measure of household net wealth (including real assets and financial assets minus debts). Taxable wealth did, however, not include pension wealth in the sense that the value of future public and occupational pensions were not included neither were savings in tax deferred pension savings accounts. Appendix B reports more details about the Swedish wealth tax.

The set of control variables we have at our disposal is quite limited, but we do have important demographics such as age of head of household, sex, family composition and marital status. We have not yet included education but will do so in the future.[8] We can condition on fixed effects, however, and that will remedy some of the shortcomings of the data.

3.2 Descriptives

Shares. Table 1 reports the percentage share of wealth tax paying households in Sweden 1968–2005. It is clear that we have information for the five top percent for most years, but complete data for the whole period are only available for the three top percent. The design of the system for taxing wealth has varied during the period, for instance concerning tax rates and exemption levels. Many more households paid wealth taxes during the 1980s and the second half of the 1990s. Almost 16 percent of the households paid the wealth tax at least once during the period. More than a third of the households that we can continuously observe 1968–2005 paid wealth taxes some time during the period.

[8]We do not observe other variables of interest, such as labor market status, occupation, health, and so on.

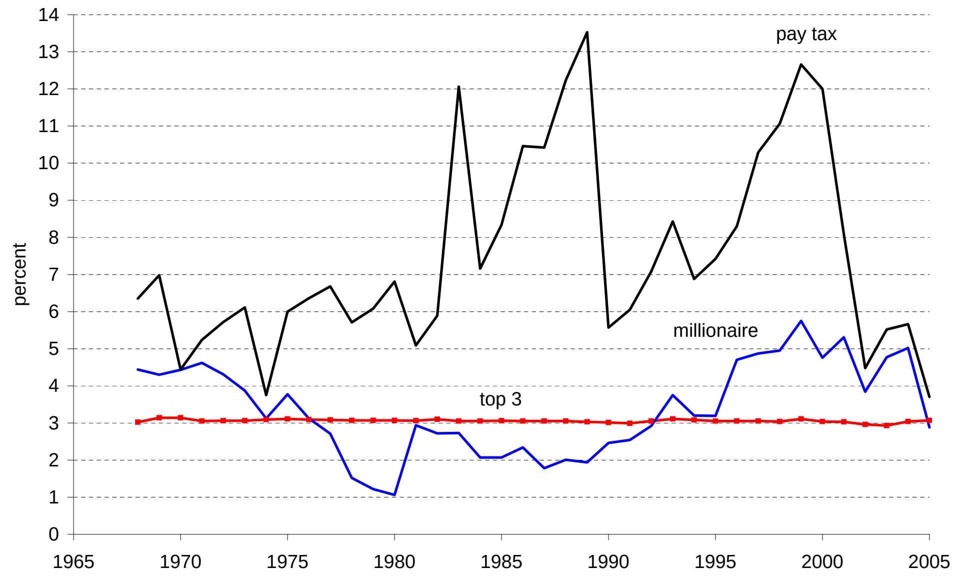

Figure 1: Shares being millionaires and paying wealth tax, 1968–2005, percent

Paying wealth tax or not is one of the possible distinctions between states that can be made for these data. Another possible distinction is between different percentiles of the wealth distribution. As mentioned above, there is only complete information over time for the *top three percent* of the wealth distribution. We will use the distinction between belonging to the top three percent or not. We can also study the flows in to and out the top three percent (across the 97th percentile, P97) and the flows within the top three percent (across P98 and P99).

Instead of this relative measure, we can also compute an absolute real measure. This will give a related but different distinction. The highest real exemption level, defined as the nominal exemption level in relation to nominal GDP per capita, during the period was the one in 1970. The real value was $\approx$ SEK_{2008} 2.25 million. This corresponds to EUR 235,000 and USD 345,000.

We have information on all fortunes above this real wealth threshold during the whole period. We will use the metaphor *millionaires* to refer to the households above this threshold. The flows of becoming a millionaire and stopping being one can also be studied.

Figure 1 reports how the share of millionaires has evolved during the period. The share of household above the real wealth threshold that we have imposed showed a decreasing trend until 1980. Since then the trend has been reversed, an increasing share of the households is above the real wealth threshold. The share above P97 is not exactly three percent when comparing two adjacent years as the panel is not balanced. The figure also shows the share paying wealth tax.

Flows and durations. Table 2 reports transitions during the period 1968–2005.

The left hand panel shows flows into and out of the top three percent of the distribution. The right hand panel reports transitions to wealth above the real wealth threshold and transitions in the reverse direction. From now on, we study transitions between two discrete states: being in the top three percent (state 1), and not being in the top percent (state 0). Alternatively, we consider being or not being a millionaire.

There is some variation over time in the inflow rates to the top three percent. This might be attributable to macroeconomic shocks and asset price changes, for instance. Most years the inflow rate is around 0.5 percent while outflow rates are in the range 15–20 percent. Obviously, inflow and outflow rates are by definition highly correlated in this case.

Turning to the second distinction, there is more variation in the inflow into being a millionaire than the inflow to the top three percent. This inflow rate is in the range 0.2–1.5 percent while the outflow rate is in the range 10–30 percent.

Mobility is closely related to duration, low mobility implies long duration. Our data give long uninterrupted accounts of wealth status. Most of our wealth spells are not censored, there are also repeated spells for some households. But there are censored spells in the beginning and the end of the period. Transitions out of the wealthy states may be because wealth has fallen below cutoff levels, but it may also be because of death or emigration.

Starting from a life cycle model perspective, we would expect it to be more likely to observe people above the cutoffs when they are in their 50s and 60s and until they retire. Transitions in to paying wealth tax, in to the top three percent, or in to becoming a millionaire would then be more likely when people accumulate wealth, while transitions in the other direction would be more likely when people have retired.

The average outflow rate from the top three percent of 17.0 suggests an average duration in the top three percent of 5.6 years. Average duration is often referred to as mean exit time (MET) in the mobility literature. The average outflow and inflow rates together imply a long run top three percent share of 3.3 percent. The actual average top three percent share is 3.05 percent.

The average outflow rate from being a millionaire of 19.17 suggests an average duration as millionaire of 5.2 years. The average outflow and inflow rates together imply a long run millionaire share of 3.3 percent. The actual average millionaire share is about the same.

While Table 2 is illustrative on average transition probabilities it masks heterogeneity in wealth transitions. Table 3 provides examples of transition paths and associated counts. The sequence `01001` means the household is in state 0 in year 1, in state 1 in year 2, back in state 0 in years 3 and 4, and in state 1 again in year 5. This households records 3 transitions. Since we have 38 years of data, the sequences are all of length 38 and start in 1968. A dot signifies a missing value. We only display sample paths of those that were continuously observed without gaps (89 percent of all cases). We only display the five most frequent patterns for a given number of transitions.

Table 2: Transitions over time, 1968–2005

	in top 3% of wealth distribution				above highest exemption threshold			
between	inflow,%	base	outflow,%	base	inflow,%	base	outflow,%	base
annual 1968–2005	0.60	6,915,775	17.70	214,116	0.65	6,900,218	19.17	229,673
1968–1969	0.48	178,541	11.68	5,558	0.54	175,927	14.73	8,172
1969–1970	0.75	179,598	21.01	5,655	1.12	177,524	19.60	7,729
1970–1971	0.52	184,913	16.41	5,815	0.88	182,524	12.15	8,204
1971–1972	0.34	192,458	9.79	6,019	0.37	189,342	13.99	9,135
1972–1973	0.37	195,112	10.75	6,092	0.32	192,641	16.22	8,563
1973–1974	0.48	197,470	13.26	6,160	0.23	195,851	24.26	7,779
1974–1975	0.67	200,588	18.59	6,251	1.09	200,525	10.99	6,314
1975–1976	0.43	203,477	11.95	6,377	0.25	202,098	22.14	7,756
1976–1977	0.44	205,799	12.83	6,461	0.25	205,748	19.61	6,512
1977–1978	1.11	207,463	33.76	6,436	0.19	208,224	49.83	5,675
1978–1979	0.57	208,582	16.40	6,476	0.15	211,860	28.14	3,198
1979–1980	0.58	209,917	16.62	6,511	0.16	213,833	24.66	2,595
1980–1981	1.07	210,231	32.82	6,535	1.99	214,513	7.10	2,253
1981–1982	0.53	207,752	15.00	6,527	0.39	208,014	19.81	6,265
1982–1983	0.72	210,992	22.31	6,593	0.66	211,809	21.50	5,776
1983–1984	0.51	208,702	14.98	6,480	0.19	209,398	29.82	5,784
1984–1985	0.50	206,636	14.35	6,398	0.35	208,687	15.37	4,347
1985–1986	0.59	203,436	17.53	6,320	0.58	205,476	12.99	4,280
1986–1987	0.66	200,495	20.03	6,240	0.24	201,948	33.28	4,787
1987–1988	0.66	195,562	17.92	5,909	0.57	198,038	14.94	3,433
1988–1989	0.58	190,812	16.78	5,810	0.37	192,806	18.89	3,816
1989–1990	0.95	188,650	28.92	5,709	0.96	190,671	19.60	3,688
1990–1991	0.65	186,731	21.79	5,792	0.60	187,774	20.70	4,749
1991–1992	0.63	185,068	18.97	5,756	0.78	185,942	15.40	4,882
1992–1993	0.58	182,199	15.74	5,678	1.12	182,444	8.01	5,433
1993–1994	0.67	182,948	19.61	5,712	0.51	181,767	25.90	6,893
1994–1995	0.48	181,912	14.62	5,672	0.51	181,702	14.21	5,882
1995–1996	0.59	179,178	18.25	5,600	1.80	178,922	6.80	5,856
1996–1997	0.49	173,135	15.70	5,464	0.88	170,172	14.60	8,427
1997–1998	0.39	170,073	12.50	5,326	0.65	166,895	10.65	8,504
1998–1999	0.65	137,789	13.09	4,040	1.47	135,162	7.51	6,667
1999–2000	0.48	150,606	13.87	4,643	0.40	146,590	21.35	8,659
2000–2001	0.66	151,058	17.93	4,535	1.38	148,502	12.37	7,091
2001–2002	0.62	150,927	19.41	4,554	0.43	147,520	32.89	7,961
2002–2003	0.45	152,080	17.44	4,736	1.28	150,671	9.59	6,145
2003–2004	0.40	153,579	12.22	4,781	0.71	150,632	10.88	7,728
2004–2005	0.83	154,327	24.65	4,778	0.26	151,205	47.11	7,900

Source: Linda, 1968-2005, full core sample.

Table 3: Individual transition patterns

# Transitions	Millionaires Pattern	# cases	Top 3% Pattern	# cases
0		333,947		337,021
	all 1's	1,714	all 1's	1,585
1		6,848		5,550
	00000000000000000000000000001111111111	108	00000000000000000000000000000000000001	307
	0000000000000000000000000000111.......	93	00000000000000000000000000000000011111	53
	0000000000000000000000000000011.......	78	00000000000000000000000000000000001111	34
	0000000000000000000000000000001.......	64	00000000000001........................	31
	00000000000000000000000000000000000001	53	00000000000000000000000000000000000111	28
	other patterns	6,452	other patterns	5,097
2		8,700		8,098
	00000000000000000000000000000001000000	189	00000000000001000000000000000000000000	62
	00000000000000000000000000000000000110	142	00000000000010000000000000000000000000	62
	00000000000000000000000000000000000010	119	00000000000000000000000000000000001000	58
	00000000000000000000000000000000010000	95	00000000000000000000001000000000000000	55
	00000000000000000000000000000000011110	94	00000000000000000000000100000000000000	54
	other patterns	8,061	other patterns	7,807
3		3,323		2,808
	00000000000000000000000001011111111111	22	00000000000000000000000000000000001001	12
	0000000000000000000000000000101.......	16	00000000000000000000000000000001100001	12
	00000000000000000000000001001111111111	15	00000000000000000000000000000000100001	7
	00000000000000000000000000001111110111	14	00000000000000000000000000000001110001	7
	0000000000000000000000000100111.......	14	00000000000000000000000000001111110001	7
	other patterns	3,242	other patterns	2,763
4–19		53,127		52,468
	series with gaps	46,168	series with gaps	46,168
all		405,945		405,945
# unique patterns (without gaps)		15,087		15,266

The Table suggests the following: (i) most households do not experience any transition, and only a tiny fraction are always rich; (ii) conditional on any movements, the number of transitions is typically small for a given individual; (iii) transitions into the higher wealth ranges occur in the second half of a series, pointing possibly to the importance of age effects.

Table 3 can only give a few (selective) examples (we record more than 15,000 different patterns in the data), so we look at summary measures of mobility next.

Mobility. An often used summary measure in the previous literature on wealth mobility is the Shorrocks' measure of mobility, see Shorrocks (1978).[9] It is defined as

$$S = \frac{N - \mathrm{tr}(P)}{N-1} \tag{7}$$

where N is the number of groups and $\mathrm{tr}(P)$ is the trace of the $N*N$ transition matrix P. The range of S is $[0, N/(N-1)]$. A higher S indicates a higher degree of mobility.

In our case, we can study four groups, each of the three top percent and those below P97 taken together. Using the average transitions rates of our data, the Shorrocks' measure is 0.386. This cannot, however, be compared to previous measures of wealth mobility in Sweden as we here only measure mobility for the top percent. The strength of our data is many observations for each household. We can, therefore, calculate a time series for annual wealth mobility for almost 40 years using the Shorrocks' measure.

Figure 2 shows how wealth mobility has evolved during the studied period. The annual Shorrocks' measures vary a lot. The figure, therefore, also includes a five year moving average. The figure suggests that wealth mobility increased during the 1970s. Wealth mobility was stable until the mid 1990s. During the last ten years the trend in wealth mobility was decreasing.

Stability. It is also possible to study wealth stability over time. Figure 3 shows the shares of household in the top three percent, respectively, that have stayed in between the percentiles where they were in the previous year. About 80 percent of the households in the top percent remained there the following year. The corresponding number for the next percent is lower. On average about 60 percent of the households between P98 and P99 remained there the following year. Stability is lower if we turn to the next percent. Slightly less than half of the households between P97 and P98 remained there the following year.

These descriptive facts certainly tell a story about how mobility in the top percent of the Swedish wealth distribution has developed during the period 1968–2005. But we have far from used all the possibilities that our panel data offer. This will be the objective of the following section.

[9]Some refer to the measure as Shorrocks' MET as it is a function of mean exit time from a group.

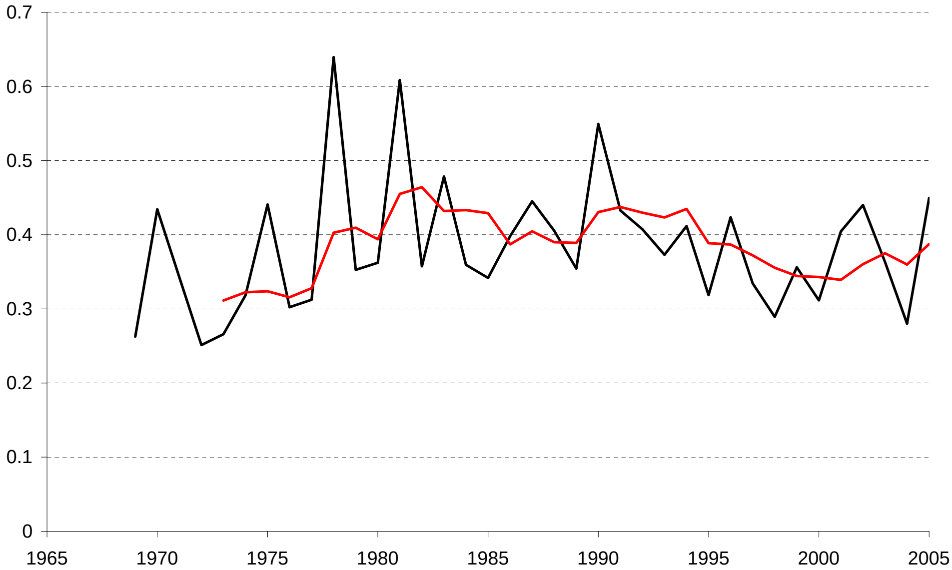

Figure 2: Wealth mobility, 1968–2005, Shorrocks' measure

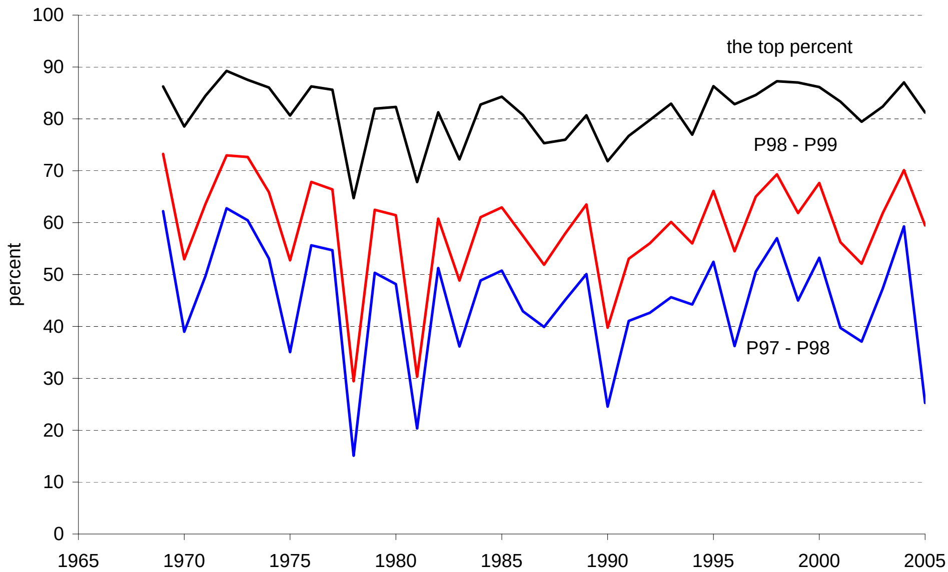

Figure 3: Wealth stability, 1968–2005, percent

4 Econometric approach

4.1 Nonlinear dynamic fixed (v random) effects panel data models

We estimate reduced form models using a dynamic specification of binary outcome models. To focus ideas, let us introduce some notation first. The simplest model is

$$y^{\star}_{it} = x_{it}\beta + \gamma y_{i,t-1} + \alpha_i + \varepsilon_{it} \tag{8}$$

where $y^{\star}$ denotes the latent variable (wealth), y is the observed outcome of interest (e.g., belong to top three percent or being a millionaire), x is a matrix of observed regressor values, α is an unobserved individual-specific, time-constant effect, and ε captures the remaining unobserved heterogeneity (error term). As usual, i and t index individuals (or households) and time, respectively. Coefficients β and γ are being estimated and are the main parameters of interest.

The distribution of wealth as observed in the data, is highly censored. Since it is practically impossible to accommodate the large degree of censoring, we abstain from any effort of modeling the continuous, but censored endogenous variable,[10] and focus on the binary outcome where the observed variable y is determined by

$$y_{it} = \mathbf{1}[y^{\star}_{it} > c_t].$$

As usual, $\mathbf{1}[A]$ is a binary 0/1 indicator taking the value one for the expression A being true, and zero otherwise.

We only observe $y^{\star}$ when it is above the threshold. Depending on model, we take c_t to be the centile corresponding to P97 of the wealth distribution in year t, or to be the value of wealth corresponding to what we call ‘a millionaire’.[11]

We shall hence estimate the probability of observing $y = 1$, conditional on regressors and the past choice for y. Of interest, then, is what Browning and Carro (2006) call ‘marginal dynamic effect’,

$$m_{it} = \Pr(y_{it} = 1|y_{i,t-1} = 1) - \Pr(y_{it} = 1|y_{i,t-1} = 0) \tag{9}$$

which can be computed with given (or estimated) values of α_i, β and γ. m_{it} tells us the difference in the probability of observing $y = 1$ depending on whether the lagged indicator is 1 or 0. The marginal dynamic effect will be zero if history does not matter in the sense that the lagged indicator does not affect the probability. The more important a lagged indicator of 1 rather than 0 is for the probability, the more positive marginal dynamic effect. Expression (9) can be computed at the *individual*

[10] In principle, suitable methods exist for moderately censored distributions: Bover and Arellano (1997) and Hu (2002) are two approaches that are applicable when the lagged endogenous variable $y^{\star}_{t-1}$ enters on the right hand side and drives dynamics, as is the case where the observability is determined by data recording, as in our case. Bover and Arellano (1997) is a random effects approach, while Hu (2002) is a fixed effects approach.

[11] Recall that this is expressed relative to nominal-per capita-deflated GDP, hence the threshold changes over time.

level, and will allow us to show the heterogeneity associated with wealth dynamics in the data.

The heterogeneity term α_i helps us to take into account time-fixed characteristics that determine the outcome but are unobserved to the analyst. In the context of our reduced-form wealth equation (that may be consistent with an equation of motion describing wealth dynamics resulting from utility maximization), one might think of preference and technological parameters that determine consumption choices (among which, prominently, the rate of time preference and the degree of risk aversion) and the process of income generation (such as worker skills, occupational trajectories and income growth parameters), both of which influence the evolution of wealth.

There are two principal ways of modeling α_i: as a fixed effect and as a random effect. For a random effects approach, the specification will need to include additional distributional parameters that have to be estimated. For the fixed effects approach, a large number of constants has to be estimated, but this can be done without distributional restrictions.[12]

In a fixed effects setting, the inclusion of α_i as individual dummy variables in maximum likelihood estimation has been viewed as causing an incidental parameter problem (Neyman and Scott, 1948). This manifests itself in inconsistent estimates of α_i, which in turn translate into biases of estimated β and γ. The estimate of an individual α_i depends on the data series $\{y_{it}, x_{it}\}_{t=1,\ldots,T_i}$, and the source of the problem is $T_i \not\to \infty$ rather than $N \not\to \infty$. The form of the asymptotic bias in the resulting estimates of β and γ can be characterized further with a specification of the distribution of ε giving rise to a particular probability model. Without further adjustments, however, finite-T estimation of a fixed effects model will not allow valid inference.

A vivid recent econometric literature investigates possibilities for estimating parameters of nonlinear (dynamic) panel data models when T is finite (and typically small in applications). See the overview in Arellano and Hahn (2007) for a synthesis of available approaches, some of which employ maximum likelihood estimates of β (and γ). One avenue relies on analytical bias reduction (Fernandez-Val, 2009; Browning and Carro, 2006; Hahn and Kuersteiner, 2004), another one on numerical jackknife methods (Hahn and Newey, 2004; Dhaene et al., 2006).

We shall in this paper, however, not employ any such methods, and instead rely on the fact that our data allow observation of long individual time series ($T_i \geq 30$), i.e., we assume that our fixed effects estimates of β and γ are not biased, and that the available series are long enough.[13] The consensus in the theoretical

[12] A logit specification for ε allows treating α_i entirely as 'nuisance' parameter such that the parameters of interest can be identified without recourse to estimates of α_i. Only in the special case of two time periods, however, is it possible to recover α_i from a closed-form solution of the first order condition of the maximum likelihood problem ('concentrating out') and to calculate (9).

[13] An open question is to what extent bias correction methods are applicable to the case of models with heterogeneous AR coefficients, as we emphasize below is insightful for the present analysis. Most methods apply to the case of a single, common, AR parameter. Browning and Carro (2006)'s

literature appears to be that the problem 'disappears' for practical purposes with T 'not small'. We explore sensitivity of our estimates by analyzing subsamples with varying lengths of T, however.

We estimate an individual dummy variable probit model, assuming an i.i.d. standard normal distribution for ε_{it}. Beyond independence, there are no distributional assumptions concerning α_i, and, in particular, they can be freely correlated with the x_{it}.

We have previously also considered the alternative of estimating a random effects model. Random effects models have the advantage of being able to generate estimates that are unbiased in the face of small T. They have obvious drawbacks, however. First, and foremost, one needs to assume orthogonality between α_i and x_{it}. This assumption is problematic in our case as we have only a very limited set of regressors that we can control for. It is likely that there will be a correlation between included regressors and the composite error $\alpha_i + \varepsilon_{it}$. Second, one needs to specify the functional form of the distribution of the individual heterogeneity. Third, calculation of (9) requires values of α_i, and a random effects approach only allows calculation of an expected marginal effect (by integrating out over the estimated distribution of α_i).

4.2 Relaxing homogeneous individual effects

There are three extensions we wish to consider. First, a simple interaction between the lagged endogenous dummy variable and the regressor matrix delivers additional flexibility, as the implied marginal effect now depends on regressor values:

$$y^{\star}_{it} = x_{it}\beta + (y_{i,t-1}z_{it})\gamma + \alpha_i + \varepsilon_{it}. \tag{10}$$

Typically, $z = x$, and (10) nests (8) if z is a vector of 1's. A simple Wald or *LR* test can be used to establish whether (10) is a statistically significant improvement over (8). In terms of interpreting the estimated equation as a dynamic wealth accumulation equation, factors that are individual-specific and impact on the return to wealth (e.g., characteristics that would influence the portfolio composition of household wealth) may be relevant, but also pure (macroeconomic) time-effects (asset price movements, for instance, that imply capital gains or losses).

Second, we can follow Browning and Carro (2006) and propose the extension

$$y^{\star}_{it} = x_{it}\beta + \gamma_i y_{i,t-1} + \alpha_i + \varepsilon_{it} \tag{11}$$

As those authors show, the simpler specification (8) implies heterogenous marginal dynamic effects despite a constant parameter γ. The added flexibility in (11) allows, on the other hand, individual dynamics and a much more insightful analysis of heterogeneous wealth dynamics, certainly under a 'fixed effects' paradigm. Also, the marginal effects distribution $\mathscr{M}$ is likely to be affected, as (11) allows it

bias corrected estimator is specifically designed for the heterogeneous case, but cannot accommodate covariates.

essentially to be 'nonparametric' while under (8) (or (10)) $\mathscr{M}$ is constrained by the probit functional form.

Third, we can consider more lags than in (11), and in theory apply the methodology to an AR(p) process. However, there are severe data limitation to be kept in mind when estimating general processes. Numerical identification requires the number of transitions observed per individual to increase with p, and, in addition, to have variation in y_t conditional on y_{t-1} at the individual level. We have succeeded to estimate at most AR(2) models with the data at hand.

We follow Greene (2004) and program a routine that allows inclusion of literally thousands of dummy variables in estimation without having to handle (compute, store, invert) associated large sparse matrices. Appendix C explains the details of the procedure for the generalization where individual processes are of the AR(p) type.

The AR(p) structure is a very flexible way of modeling individual dynamics, certainly compared to what else has been done in the relevant literature. Yet, this modeling approach does not allow for moving average components, as would be the case with more general models (of the ARMA(p,q) type). Extending the current methodology based on maximum likelihood to such general error structures appears practically infeasible, however.[14]

4.3 Interpreting heterogeneity

Browning and Carro (2007) warn that 'heterogeneity is too important to be left to the statisticians'. While it is conceptually easy to add in heterogeneity in parameters as we do in this paper, the question is, whether there is much of an interpretation to these parameters, or whether they are, as often, viewed as nuisance parameters that one may want to control for, but are not of prime interest in themselves.

Since our focus is on wealth mobility, we are in a position, however, to give the estimated marginal dynamic effect (9), obtained from an estimate of γ, meaningful interpretation. In particular, Shorrocks' measure of mobility (7), that is based on the off-diagonal elements of the transition matrix, is directly linked to this magnitude.

Denote by $P_{ij} = \Pr(y_t = j|y_{t-1} = i)$ the elements of the transition matrix. For binary outcome transitions ($i, j = 0, 1; N = 2$), S equals

$$S = 1 - P_{00} + 1 - P_{11}$$

The marginal dynamic effect that we estimate equals $m = P_{11} - P_{01}$. With $P_{00} + P_{01} = 1$, we have $S = 1 - m$.

[14] Some attempts to explore alternative avenues for simpler, univariate time series data have been undertaken: Poirier and Ruud (1988) consider estimation of ARMA probit models (and other general error structures) by what they call generalized conditional moment estimation, a variant of GMM. Gourieroux et al. (1984) propose a related method. As one alternative, Browning et al. (2009) propose a simulated minimum distance (indirect inference) estimator applied to income dynamics. This is computationally intensive, however, unlike our approach which is very fast.

5 Empirical evidence

5.1 Data and sample selection

We study two separate models. The first dependent variable of interest is 'being in the top three percent' of the wealth distribution, which is a year-specific indicator that has value one if the household belongs to the wealthiest three percent in any given year, using all available observations. The second dependent variable is a year-specific indicator taking the value one if the household has real wealth above the real wealth threshold, i.e., the household is a 'millionaire'.

We start with the full core sample, as described earlier, at the household level. All heads of household are 18 and above. There are more than 400,000 different households and more than 7.5 million observations. We only select those households where the identity of the head of household does not change over time, to avoid problems in interpreting coefficients. This leaves about 350,000 households. Not doing so would introduce additional variation in the regressor values that are taken to be head-of-household characteristics.[15]

We further throw out observations with missing values in regressors, and remove all households that never experience any transitions in the dependent variable. For both dependent variables that we consider, this leaves about 25,000 households with more than 600,000 observations (the 'millionaires' sample is slightly larger).

Further, we only select those with heads of households born in the period 1940–1950. There are about 6,000 such households. This is to make sure that we have a reasonably large, but also reasonably homogeneous cohort that entered the labor market towards the beginning of the observation period, and has not yet retired by the end of it (i.e., these heads of households are between 18 and 28 in 1968 and between 55 and 65 in 2005). Retirement brings about the issue of accounting for wealth decumulation, and perhaps it is a good idea to first leave this out of the picture.

We then select only those who have at least 30 observations over time, so as to be reasonably sure that the fixed effects approach is okay. There are about 2,000 households left.

We further need to have enough variation in the dependent variable and its lag in order to estimate heterogeneous AR models as in (11). This AR(1) specification will be our baseline model. We are left with 980 households, good for almost 34,000 observations across all years, when studying transitions into and out of the top three percent. The sample for transitions into and out of being a millionaire is a bit larger – 1,303 households and more than 45,000 observations.

We model the probability of being in the top of the distribution and being a millionaire as functions of the following groups of covariates that we have at our

[15] The head of household is selected as follows: each person has a person specific identifier and a household identifier. The data come in such a format that the household identifier equals the person identifier of one of its members. We take that particular member to be head of household.

disposal:

- head-of-household characteristics; age, decade of birth cohort, gender, and marital status
- household demographics; number of children in different age ranges, and household size
- macroeconomic factors; real GDP growth including an interaction with an indicator for negative growth; growth in the stock market index
- wealth tax policy parameters: exemption levels depending on family composition, and the marginal tax rate applicable to the first tax bracket

Since the underlying model takes into account unobserved heterogeneity by means of fixed effects, we, as it were, remove cohort effects and gender differences without controlling for them by means of variables. We presently do not observe the head of household's education, but such effects would also be mainly accounted for in the fixed effects estimates.[16]

We have, without presenting them here, estimated static models such as fixed and random effects logit models, but also static versions of the fixed effects probit model. From these estimates we learn two things: first, a standard Hausman specification test clearly rejects the random effects specification. This is one important reason to prefer the fixed effects approach. Second, we find in fact very similar parameter estimates between the conditional logit model and the dummy variable probit (except for the fact that the scale is different). This is no evidence for the absence of a bias in the estimated γ coefficient, but at least reassuring of no major impacts of the ancillary parameter estimates on the βs.

Presently we abstain from a detailed discussion of parameter estimates but focus on distributions of heterogeneous parameters.

We now present a number of empirical models that were estimated using the data. They enable us to keep things constant that implicitly varied in the pictures and table discussed before. These are all reduced form models, but they may be interesting in their own right.

5.2 Belong to the top three percent

We now turn to simple dynamic specifications where we have put in the lagged value of being in the top three percent as regressor. The model is that of equation (8), with individual intercepts and and a common AR(1)-parameter. Parameter estimates are in the first column in Table 4.

We start with model (8). The overall fit of the model is relatively good, as calculated from weighted prediction/realization tables (we predict $\Pr(x_{it}\beta + \eta_i)$ and assign the value of 1 if larger than 0.5, and subsequently calculate a prediction score). The value is 56.3 percent.

[16] Income is not readily available at the moment, but we aspire to construct income series as well.

Table 4: Dynamic fixed effects probit models for being in the top 3% of the wealth distribution

	single AR(1) parameter		AR(1) parameter interacted with regressors		heterogeneous AR(1) parameters	
	coeffi-cient	standard error	coeffi-cient	standard error	coeffi-cient	standard error
lagged endogenous	1.682815	.0214667***	7.340641	2.580276***		
age	-.159051	.0748251**	-.2482446	.0971946**	-.1997144	.0821888**
$age^2/10$	.0527786	.0169562***	.0801583	.0221066***	.0641682	.0186399***
$age^3/100$	-.0043122	.0012401***	-.0067292	.0016245***	-.0052163	.0013644***
cohabiting	.0323854	.2051512	-.0631404	.2535309	-.0330839	.2260541
widow/er	-.0710483	.1172976	-.0927725	.1440923	-.1123935	.1257597
divorcee	.4490564	.1802275**	.5310524	.2060472***	.5180835	.1926504***
single	-.0014392	.1301047	.0121976	.1573418	.0125043	.1399965
lone parent	.484653	.1231247***	.4064372	.1478057***	.4659247	.1363659***
#kids<18	-.2161894	.1319195	-.2754027	.1633705*	-.1820396	.1420796
HH size	.4132108	.1017754***	.3601143	.1285108***	.4249284	.1083338***
#kids × age	-.0016641	.0019025	.000972	.0022904	-.0024679	.0020744
real GDP growth	-.0285804	.0106788***	-.0670244	.0136515***	-.0325305	.0111584***
(0/1) growth < 0	.0973162	.0613178	.1386106	.071006*	.1063557	.0636562*
real GDP gr if <0	.1187836	.0383199***	.2871058	.0456194***	.1358842	.0398084***
growth stock mkt index	.108984	.0506212	.0240268	.0645019	.1330744	.053012**
marginal tax rate	.0759754	.0419888*	.1166482	.0507147**	.0774579	.0452411*
(indiv.) exemp. level	.0123594	.0467189	.1110421	.0584277*	.0140471	.0491688
interaction terms y_lag						
age			-.2520776	.1742445		
$age^2/10$			.0266211	.0388981		
$age^3/100$			-.0002289	.0028123		
cohabiting			.1438271	.3523482		
widow/er			.103931	.2277467		
divorcee			-.0781973	.3058134		
single			.0360356	.2298273		
lone parent			.2435017	.2455084		
#kids<18			.3441656	.2748643		
HH size			.1994754	.2104067		
#kids × age			-.0120686	.0037633***		
real GDP growth			.1163228	.0232157***		
(0/1) growth < 0			-.2192608	.1464162		
real GDP gr if <0			-.5423698	.0920906***		
growth stock mkt index			.124521	.1095409		
marginal tax rate			-.0296074	.1344551		
(indiv.) exemp. level			-.2463421	.0987869***		
log-likelihood	-10307.47		-10167.15		-9867.90	
	Number of obs = 33787		Number of HH = 980		Number of times per HH = 31-38	

Compared to a static model estimated on the same sample (not displayed), the dynamic specification is a substantial improvement. The *LR*-test statistic is 6,626.0 at 1 degree of freedom.

The lagged endogenous variable is, therefore, statistically very important, but a more interesting magnitude to look at from an economic point of view is the implied marginal dynamic effect (9), since it characterizes individual wealth mobility. This marginal dynamic effect is household-specific since it is predicted from $\Phi(\bar{x}_{it}\beta + \gamma y_{i,t-1} + \eta_i)$, i.e., even though γ is constant, the probability changes between individuals due to the household specific constant η_i.

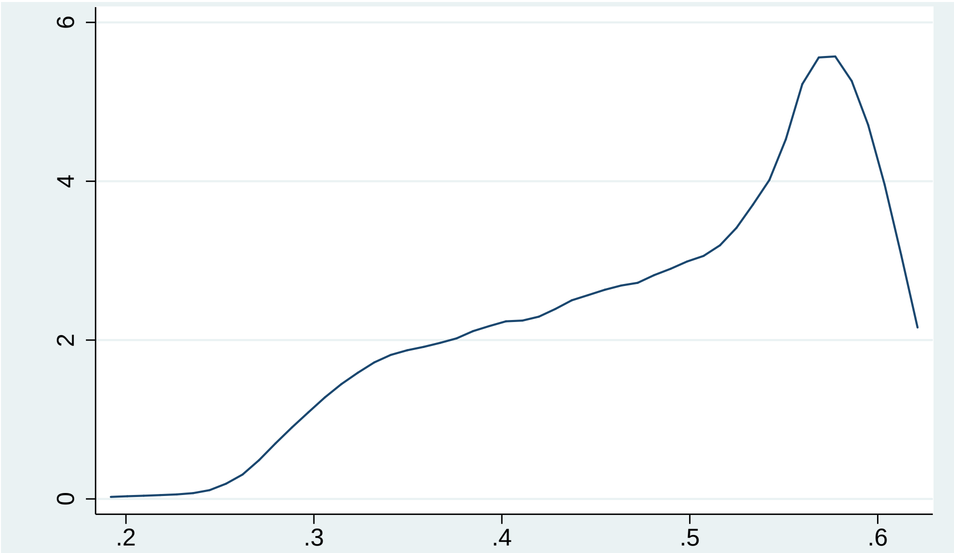

Figure 4: PDF of predicted marginal dynamic effect, FE probit, common AR parameter

Figure 4 shows the pdf of these marginal dynamic effects. It is clear that there is a large concentration at the upper values. The AR-effects appear to be rather large at 0.25–0.6 for most observations.

The flexibility of the model can be enhanced by considering specification (10). Parameter estimates are in the second column of Table 4. A *LR* test again suggests a substantial improvement. The implied marginal dynamic effect distribution is, however, not so much affected, as can be gleaned from Figure 5.

We now turn to estimates from a more flexible specification. The model is (11), with individual intercepts and AR(1)-parameters. Parameter estimates are in the third column in Table 4. Comparison with the restricted model from above suggests that there are some, but no dramatic changes in coefficient estimates. The flexible model is also a statistical improvement over (8) according to an *LR* test, even though the overall fit is not much improved (the prediction score changes to 56.8 percent).

Note that the differences between these three specifications in terms of coefficient estimates of the included regressor variables is, actually, minor. The age function does display a life-cycle pattern with strong accumulation in the mid-age range, slowing down until shortly before retirement age. We abstain from a detailed discussion of other parameters at the moment.

Many of the 980 additional AR(1) parameters are statistically significant; 440 have a significance level of 1 percent or lower, 227 of between 1 and 5 percent, and another 75 of between 5 and 10 percent. The marginal effects distribution is very different, however, compared to the restricted model, see Figure 6. In particular, the domain of the distribution is much more spread out, and even a tiny fraction

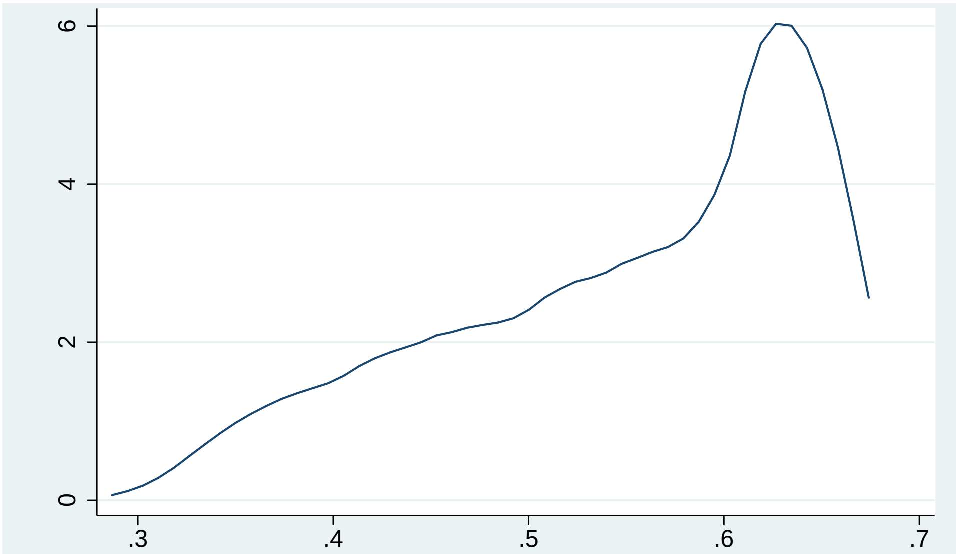

Figure 5: PDF of predicted marginal dynamic effect, FE probit, common interacted AR parameter

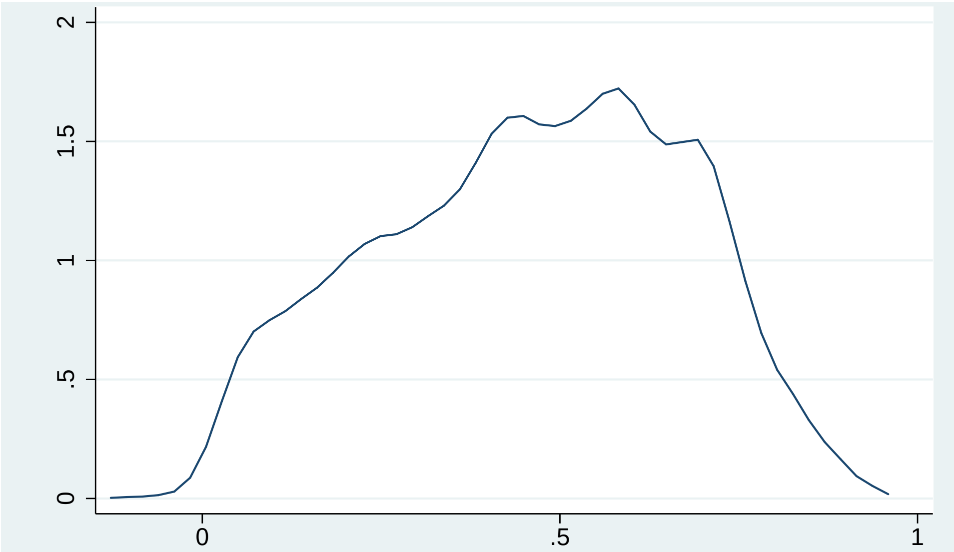

Figure 6: PDF of Predicted Marginal Dynamic Effect, FE Probit, individual AR parameter

of negative marginal effects are observed.[17] That is, there is substantial individual heterogeneity in wealth trajectories that the simpler model fails to capture.

We can, in principle, allow for more flexible time series patterns by extending the lag order of the AR model. This implies a significant drop in observations, however, since higher lags call for more observed transitions at the individual level. In order not to lose so many observations, we can instead use the current sample and estimate an AR(2) model for those individuals that have a sufficient variation, and we impose an AR(2) coefficient of zero on the remaining individuals. We refer to this model as the 'mixed model'.

Conducting this exercise leads to the following conclusions: the mixed model is not a statistical improvement over the AR(1) model. There are only 18 households with an AR(2) coefficient that has a significance level of 10 percent or lower (only 2 at 1 percent or lower). The associated likelihood ratio test also rejects the mixed model as a statistical improvement over the AR(1) model. We, hence, conclude that a flexible AR(1) model is a good empirical description of wealth transitions for the data at hand.

We now have a short look at the correlates of our measured heterogeneity. We regress the estimated individual parameters on observable characteristics, plus year of birth indicators and a gender dummy. We use time-averaged values of the other regressors. The first estimation reported in Table 5 asks the questions what the correlates of the fixed effects are. We do not find many strong correlations with regressors.

[17]Unlike Figures 4 and 5, Figure 6 is based on predictions at individual regressor values, but that as such is immaterial.

The second pair of columns in Table 5 extends the analysis to the estimated individual AR(1) parameters (γ_i). Looking at the coefficient estimates suggests that a few variables are correlated with individual wealth dynamics. Both household size and gender appear important. Lastly in the third pair of columns of Table 5 we can look at the correlates of the marginal dynamic effects whose distribution was displayed earlier. Here, cohort effects appear to play some role. In general, however, it appears difficult to 'explain' these individual specific parameters from observables. This suggests that an alternative modeling strategy whereby full interaction of all observables were used to capture heterogeneity would not be able to explain as much of the variation in wealth transitions.

Table 5: Correlates: Top 3%

	Fixed Effects		AR(1) parameters		Marginal Dynamic Effects	
	coefficient	standard error	coefficient	standard error	coefficient	standard error
age†	2.118387	3.153498	-1.050247	3.582116	.004712	1.14794
(age†)^2/10	-.5240023	.7325195	.2440081	.8320821	-.0052826	.2666526
(age†)^3/100	.0420609	.0565761	-.0180836	.0642658	.0009396	.0205949
cohabiting†	-.3910897	.6489678	-.6514766	.7371743	-.2411752	.2362381
widow/er†	-.040489	.7992894	1.322205	.9079273	.5157881	.2909583*
divorcee†	-.9447946	.7997925	1.148084	.9084987	.3148	.2911414
single†	-.0204594	.8616521	.8122112	.9787663	.3627672	.3136597
lone parent†	-.498079	.470098	-.4379938	.5339929	-.1175987	.1711257
(#kids<18)†	.4873501	.8192932	-.4284808	.93065	.0414967	.2982401
HH size†	-1.14095	.6310005*	1.637889	.7167649**	.4187195	.2296976*
#kids × age†	.0118689	.0127949	-.0271198	.014534	-.0107593	.0046576**
(indiv.) exemp. level†	1.076342	1.02917	-.7688777	1.169053	.0187085	.3746397
female	.2012335	.1190079*	.4425653	.1351833***	.0825061	.0433214*
born 1941	.0000376	.0721246	.0593493	.0819276	.012235	.0262549
born 1942	-.0032152	.0758997	.1380168	.0862158	.0474163	.0276291*
born 1943	.0241302	.0811675	.0971048	.0921996	.0343098	.0295467
born 1944	-.0416399	.0847683	.1129717	.0962898	.0401895	.0308574
born 1945	.0362068	.0941263	.0995891	.1069198	.0393375	.034264
born 1946	-.0502707	.1023095	.1506672	.1162152	.0635199	.0372428*
born 1947	.0530385	.1104681	.0786957	.1254827	.0199092	.0402127
born 1948	-.0523061	.1276082	.1774652	.1449525	.0692565	.0464521
born 1949	-.0454891	.1412671	.035303	.1604679	.0185861	.0514242
born 1950	.0771499	.1956887	-.0308003	.2222864	-.0019476	.0712348
constant	-28.24816	45.12558	13.71749	51.25897	-.4015446	16.42667
Adj. R^2	0.1625		0.0074		0.0097	
†averages over time			Number of HH = 980			

Note: This Table shows regression results from regressing estimated constants and AR(1) parameters as well as marginal dynamic effects on household (average) regressors. *, **, and *** denote statistical significance at the 10, 5, and 1 percent level, respectively.

5.3 Be a millionaire

So far, we looked at transitions within the wealth distribution, relative to other households. This subsection now analyzes wealth mobility in an absolute sense: which households become 'millionaires'?

Parameter estimates for the simple dynamic model (8) are in the first set of columns of Table 6. Abstaining from a detailed discussion of coefficient estimates, we confine ourselves to remarking the similarity in conclusions compared to the 'top-3%' specification from above, and the similarity of coefficient estimates displayed in the Table across specifications.

We may want to note that, as before, the model is a statistical improvement over a static model estimated on the same sample. We also find this model to provide a good fit to the data. We have calculated marginal effects, and find, without displaying a figure here, that the distribution resembles its counterpart for the top three percent exercise before (Figure 4).

Similar remarks apply to the more flexible model (10). Again, we end up preferring a specification with one individual AR(1) parameter, see the third pair of columns of Table 6. Among the 1,303 estimated values, 349 have a significance level of 1 percent or lower, 351 of 1–5 percent, and 163 are significant at a level of 5–10 percent.

Recall that there is a direct, inverse relation between the marginal dynamic effect and Shorrocks' mobility index (subsection 4.3). We may conjecture that age and demographics are important determinants of life cycle mobility. For this reason, we also estimated the same model without age and demographics (*LR* test statistic of 2,976.3 at 11 degrees of freedom). Figure 7 now displays the distribution of individual mobility indexes for both models. It is clear that not accounting for demographics and age gives a very misleading picture of mobility indeed, as in that case much of the observed life cycle mobility is attributed to the individual-specific parameters. We would erroneously characterize the society as substantially less mobile.

Table 7 regresses the various individual parameters on (time-averaged) regressor values. We find a few more significant correlates here. Dynamics correlate with household size, marital status, and year of birth cohorts. We abstain from detailed discussion, though.

Similar to the top three percent case, we find also in the millionaires sample only a handful of observations with a significant AR(2) parameter, once we consider the appropriate extension. The AR(1) model is our preferred model also in this case.

5.4 Final remark on time series dimension

We rely, as indicated, on the availability of sufficiently long individual time series that are long enough as to be able to virtually ignore the incidental parameters problem. There are no established values of T that allow a clear-cut classification into what may be considered 'long' (as opposed to 'short'). Available Monte Carlo evidence is specific to set-ups and may not carry over to our application, but we stay far above the values typically considered for T.[18] We select a sample in which

[18]Heckman (1981) considers $T \geq 8$ reasonably long enough, but his results are questioned by Greene (2004) whose analyses suggests values in excess of $T > 16$ to be needed (also see: Browning

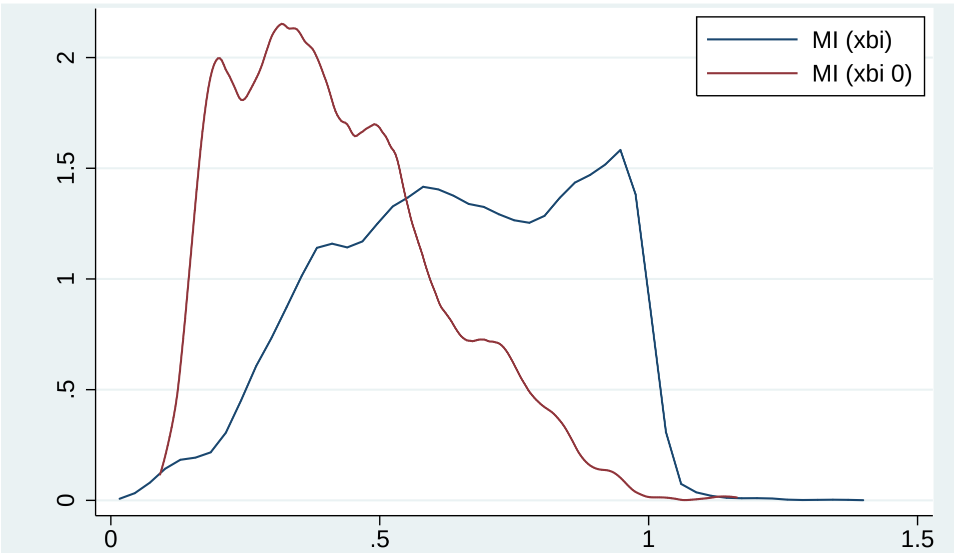

Figure 7: PDF of Predicted Mobility Index, 'Millionaires', FE Probit, individual AR parameter (MI(xbi)–mobility index calculated per individual and predicted from demographics and age, MI(xbi 0)–mobility index calculated per individual without demographics or age).

Table 6: Dynamic fixed effects probit models for being a millionaire

	single AR(1) parameter		AR(1) parameter interacted with regressors		heterogeneous AR(1) parameters	
	coeffi-cient	standard error	coeffi-cient	standard error	coeffi-cient	standard error
lagged endogenous	1.542982	.0204381***	12.81794	2.403871***		
age	-1.016732	.064382***	-.9576416	.0851081***	-1.180487	.0716111***
$age^2/10$	.247407	.0146571***	.2338048	.0195058***	.2867058	.0163709***
$age^3/100$	-.0181976	.0010733***	-.0169749	.0014392***	-.0210695	.0012028***
cohabiting	.0293042	.1946713	-.0074937	.2253558	.003391	.2112389
widow/er	-.0473421	.1099141	-.2188473	.1387833	-.0789471	.1183841
divorcee	.3002816	.1706408*	.1135587	.2076083	.3554883	.1875367*
single	-.0656721	.1232698	-.1712934	.1530586	-.0900556	.1340413
lone parent	.5311998	.1170961***	.6031929	.1408785***	.549723	.129328***
#kids<18	-.0115711	.1253915	-.0168296	.1602677	.0524657	.136399
HH size	.445204	.0945759***	.3835701	.1230725***	.467978	.1008531***
#kids × age	-.0082216	.0018199***	-.0068419	.0022401***	-.0100151	.002014***
real GDP growth	-.1102544	.0099726***	-.1787547	.0131464***	-.1172893	.0105225***
(0/1) growth < 0	.0261611	.0578067	.1349433	.0644732**	.0961051	.0601072
real GDP gr if <0	.2708629	.036116***	.5431842	.0415277***	.3246629	.0374685***
growth stock mkt index	.1828041	.0477262***	.154156	.0620102**	.2649956	.0504448***
marginal tax rate	.0256652	.0442615	.1697491	.055845***	.0532411	.047151
(indiv.) exemp. level	.1985484	.0429015***	.1602001	.0531052***	.1953812	.0455879***
interaction terms y_lag						
age			-.6705865	.1612835***		
$age^2/10$			.1311436	.0356409***		
$age^3/100$			-.0087771	.002547***		
cohabiting			-.0426035	.2834634		
widow/er			.3871499	.2100973*		
divorcee			.4366429	.270189		
single			.2405958	.2102645		
lone parent			-.1773855	.2294394		
#kids<18			.2018143	.2608769		
HH size			.2578339	.1938057		
#kids × age			-.0088382	.003627**		
real GDP growth			.1901901	.0219909***		
(0/1) growth < 0			-.7336487	.1776141***		
real GDP gr if <0			-1.073245	.112499***		
growth stock mkt index			.028123	.104533		
marginal tax rate			-.4411231	.1103541***		
(indiv.) exemp. level			-.0540775	.0928149		
log-likelihood	-12184.25		-11889.69		-11565.47	
	Number of obs = 45066		Number of HH = 1303		Number of times per HH = 31-38	

the estimated marginal dynamic effects distribution does not change appreciably

and Carro (2007)). At the other extreme, Browning and Carro (2006) feel very safe with ignoring bias with micro data that have $T \geq 100$ (weekly observations), but unfortunately, they do not report what happens when they choose shorter panels. Even bias correction methods need 'long enough' series to deliver satisfactory results; Hahn and Kuersteiner (2004) consider $T = 16$ 'moderately large'; Fernandez-Val (2009) reports $T = 8$ to work and considers at most $T = 16$; Browning and Carro (2006) settle on $T = 9$.

Table 7: Correlates: Millionaires

	Fixed Effects		AR(1) parameters		Marginal Dynamic Effects	
	coeffi-cient	standard error	coeffi-cient	standard error	coeffi-cient	standard error
age†	8.142965	3.555182**	.9568186	3.849714	1.345923	1.209481
$(\text{age}^{\dagger})^2/10$	-1.933486	.8213592**	-.2792609	.8894055	-.3361594	.2794283
$(\text{age}^{\dagger})^3/100$	.1513464	.0631119**	.0260597	.0683405	.0276565	.0214708
cohabiting†	-.7290283	.5600822	-.4781342	.6064828	-.2816312	.1905413
widow/er†	-.0888734	.7888198	2.814582	.8541703***	.968578	.2683583***
divorcee†	-.8222488	.7966586	2.984527	.8626586***	.900239	.2710251***
single†	-.0281916	.8488197	2.736601	.919141***	.9162401	.2887704***
lone parent†	-.3667024	.4603536	-.4907778	.4984921	-.1201323	.1566134
(#kids<18)†	1.002271	.8072598	-1.327649	.8741381	-.1827494	.2746316
HH size†	-1.172323	.6059923*	1.643329	.6561964**	.3816931	.20616*
#kids × age†	.0030055	.0128408	-.0053662	.0139046	-.0038386	.0043685
(indiv.) exemp. level†	.7802054	.9627951	1.800913	1.042559	.7706545	.3275451**
female	.1089658	.1144269	.2005417	.1239067	.0053418	.0389283
born 1941	.0059612	.0703328	.2078861	.0761596***	.0538252	.0239274**
born 1942	-.0253665	.0741187	.1990982	.0802591**	.0427885	.0252154*
born 1943	.0623988	.0790037	.2120676	.0855489**	.0700402	.0268773***
born 1944	-.0262492	.0844928	.1736	.0914927*	.0442154	.0287446
born 1945	-.0248961	.0935969	.1130913	.101351	.0264545	.0318419
born 1946	-.0580893	.1015184	.1551799	.1099287	.0521474	.0345368
born 1947	-.0502479	.1115674	.0401641	.1208104	-.0175432	.0379555
born 1948	-.1233472	.1273414	-.0245195	.1378912	-.0157487	.0433218
born 1949	-.100163	.1415697	.0629306	.1532982	.0015334	.0481623
born 1950	-.0719103	.1940168	.1412618	.2100904	.0294464	.066005
constant	-99.97246	51.21036*	-14.66973	55.45294	-19.20964	17.42188
Adj. R^2	0.1567		0.0126		0.0378	
†averages over time			Number of HH = 1303			

Note: This Table shows regression results from regressing estimated constants and AR(1) parameters as well as marginal dynamic effects on household (average) regressors. *, **, and *** denote statistical significance at the 10, 5, and 1 percent level, respectively.

anymore, and find $T = 30$ to be a reasonable lower value.[19,20]

6 Conclusions

With increasing availability of suitable micro data, the recent economic literature has seen a surge in interest in studying distributional issues and implications of top incomes. In addition, there is revived interest in studies on wealth mobility.

We add to this literature the aspect of studying individual wealth mobility over

[19] Figures A.1 and A.2 in the Appendix display the outcome of the following exercise: we start with a sample with $T = 37$ (the maximum in the data, after allowing for an initial value), and randomly eliminate time periods from individual series, successively reducing T. This heuristic approach, while debatable (it creates gaps in individual time series that are ignored), shows a large degree of stability in the distribution for all $T \geq 30$. Only below this value, do we see first changes. These figures are not congruent with Figure 6 due to differences in the base sample.

[20] We have also estimated (8) under the restriction $\gamma = 0$, and found no large differences in parameter estimates of β between our fixed effects probit and a conditional logit (save for differences in scale, of course) with our choice of T.

the entire working life cycle, exploiting long individual time series of household wealth.

We use a large administrative sample from Sweden. The period under study covers the years 1968–2005. We can track many households that are continuously in the sample.

The wealth data are heavily censored from below, owing to the fact that their values originate from wealth tax registers that only keep wealth records for actual taxpayers. The wealth tax in Sweden (abolished as per 2007) was associated with relatively high exemption levels. This leaves only a small fraction (between 3.4 and 13.1 percent) of households observed with wealth in any one cross section. However, we capture the top of the wealth distribution, which is very important for determining macroeconomic aggregates. And from a life cycle point of view, there is actually a large fraction of households (34 percent) that ever pay wealth taxes at some point during their life cycles if we condition on those that are in the sample every year from 1968 to 2005.

Whereas the wealth information available in the tax data is restricted, we can shed new light on the study of wealth mobility and dynamics at the individual level. Due to heavy censoring, we confine ourselves to looking at changes over time in binary indicators. We, thus, study movements in and out of the top three percent of the wealth distribution, and across an absolute wealth threshold that we refer to as 'millionaires' (in local currency). We rely on a fixed effects approach. Owing to the comparatively large time dimension of our data, we estimate dynamic binary choice models that allow calculating a fixed effect at the individual level, and importantly, that allow for estimating parameters of individual wealth dynamics. We present evidence based on AR(1) processes. We test for and reject presence of higher-order lags in individual time series, a feature of our study that is unique in the empirical literature on wealth.

We can also show that simpler specifications without accounting for the substantial heterogeneity can lead to a misleading characterization of wealth accumulation behavior over the course of a households' working life cycle. The evidence we provide in this paper is principally informative for macroeconomic modeling of life cycle consumption models that rely on individual heterogeneity of one form or other to pin down unobserved parameters by way of calibrating to the cross sectional wealth distribution. Sample paths of simulated models have been studied in the literature. Empirical evidence on actual household wealth transitions over long time spans has been lacking heretofore.

References

Arellano, M. and Hahn, J. (2007). Understanding bias in nonlinear panel models: Some recent developments. In Blundell, R., Newey, W., and Persson, T., editors, *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, volume 3. Cambridge University Press, Cambridge.

Bertola, G. (2000). Macroeconomics of distribution and growth. In Atkinson, A. B. and Bourguignon, F., editors, *Handbook of Income Distribution*, volume 1, pages 477–540. Elsevier, Amsterdam.

Bover, O. and Arellano, M. (1997). Estimating dynamic limited dependent variable models from panel data. *Investigaciones Economicas*, 21(2):141–165.

Browning, M. and Carro, J. (2006). Heterogeneity in dynamic discrete choice models. Discussion Paper 287, Department of Economics, Oxford University.

Browning, M. and Carro, J. (2007). Heterogeneity and microeconometrics modelling. In Blundell, R., Newey, W., and Persson, T., editors, *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, volume 3, pages 47–74. Cambridge University Press, Cambridge.

Browning, M., Ejrnæs, M., and Alvarez, J. (2009). Modeling income processes with lots of heterogeneity. unpublished, Department of Economics, Oxford University.

Browning, M., Hansen, L. P., and Heckman, J. (1999). Micro data and general equilibrium models. In Taylor, J. and Woodford, M., editors, *Handbook of Macroeconomics*, volume 1A, pages 543–633. Elsevier, Amsterdam.

Cagetti, M. and De Nardi, M. (2009). Estate taxation, entrepreneurship, and wealth. *American Economic Review*, 99:85–111.

Castañeda, A., Díaz-Giménez, J., and Ríos-Rull, J.-V. (2003). Accounting for the U.S. earnings and wealth inequality. *Journal of Political Economy*, 111(4):818–857.

Davies, J. B. and Shorrocks, A. F. (1999). The distribution of wealth. In Atkinson, A. B. and Bourguignon, F., editors, *Handbook of Income Distribution*, pages 605–675. Elsevier, Amsterdam.

Dhaene, G., Jochmans, K., and Thuysbaert, B. (2006). Jackknife bias reduction for nonlinear dynamic panel data models with fixed effects. unpublished, Department of Economics, K.U. Leuven University.

Dynan, K., Skinner, J., and Zeldes, S. (2004). Do the rich save more? *Journal of Political Economy*, 112(2):397–444.

Edin, P.-A. and Fredriksson, P. (2000). LINDA - Longitudinal INdividual DAta for Sweden. Working Paper 2000:19, Department of Economics, Uppsala University.

Fernandez-Val, I. (2009). Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics*. forthcoming.

Gourieroux, C., Monfort, A., and Trognon, A. (1984). Estimation and test in probit models with serial correlation. In Florens, J., Mouchart, M., Raoult, J., and Simar, L., editors, *Alternative Approaches to Time Series Analysis*. Publications des Facultes Universitaires Saint-Louis., Brussels.

Greene, W. H. (2004). The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *Econometrics Journal*, 7:98–119.

Hahn, J. and Kuersteiner, G. (2004). Bias reduction for dynamic nonlinear panel models with fixed effects. unpublished, Department of Economics, UCLA.

Hahn, J. and Newey, W. K. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72:1295–1319.

Heckman, J. J. (1981). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In Manski, C. F. and McFadden, D., editors, *Structural Analysis of Discrete Data with Econometric Applications*, pages 179–197. MIT Press, Cambridge.

Hu, L. (2002). Estimation of a censored dynamic panel data model. *Econometrica*, 70(6):2499–2517.

Huggett, M. (1996). Wealth distribution in life-cycle economies. *Journal of Monetary Economics*, 38(3):469–494.

Hurst, E., Luoh, M. C., and Stafford, F. P. (1998). Wealth dynamics of American families, 1984–94. *Brookings Papers on Economic Activity*, (1):267–337.

Jappelli, T. and Pistaferri, L. (2000). The dynamics of household wealth accumulation in Italy. *Fiscal Studies*, 21(2):269–295.

Jianakoplos, N. A. and Menchik, P. L. (1997). Wealth mobility. *Review of Economics and Statistics*, 79(1):18–31.

Keister, L. A. (2005). *Getting Rich: America's New Rich and How They Got That Way*. Cambridge University Press, Cambridge and New York.

Klevmarken, N. A. (2004). On the wealth dynamics of Swedish families, 1984–98. *Review of Income and Wealth*, 50:469–491.

Klevmarken, N. A., Lupton, J. P., and Stafford, F. P. (2003). Wealth dynamics in the 1980s and 1990s: Sweden and the United States. *Journal of Human Resources*, 38:322–353.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16(1):1–32.

Poirier, D. J. and Ruud, P. A. (1988). Probit with dependent observations. *Review of Economic Studies*, 55:593–614.

Prentice, R. L. and Gloeckler, L. (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*, 34(1):57–67.

Selin, H. (2009). The rise in female employment and the role of tax incentive: An empirical analysis of the Swedish individual tax reform of 1971. Working Paper 2009:3, Uppsala Center for Fiscal Studies, Department of Economics, Uppsala University.

Shorrocks, A. F. (1978). The measurement of mobility. *Econometrica*, 46:1013–1024.

Steckel, R. H. and Krishnan, J. (2006). The wealth mobility of men and women during the 1960s and 1970s. *Review of Income and Wealth*, 52(2):189–212.

Appendix A Additional tables and figures

Table A.1: Total sample sizes, by sample and year

	P-Sample		F-Sample		percentage	percentage
year	number of households	number of individuals	number of families	number of individuals	identical overlap (P:F)	common individuals
1968	263,605	640,202				
1969	268,925	649,542				
1970	271,204	629,069				
1971	269,329	710,090				
1972	269,376	709,722				
1973	269,991	697,714				
1974	271,800	705,580				
1975	273,988	703,978				
1976	274,634	700,985				
1977	275,601	700,099				
1978	273,931	693,697				
1979	275,054	692,866				
1980	275,872	688,910				
1981	276,240	685,804				
1982	281,135	687,341				
1983	279,629	675,478				
1984	281,879	677,777				
1985	282,945	673,981				
1986	284,242	670,428				
1987	285,519	669,943				
1988	288,547	671,950				
1989	291,958	680,026				
1990	291,644	693,590				
1991	288,044	672,857	285,102	772,253	65.0	98.1
1992	289,368	671,742	286,860	777,973	64.4	98.2
1993	291,068	696,753	289,022	784,065	67.6	98.4
1994	293,130	698,601	291,396	790,005	67.9	98.7
1995	294,086	698,513	292,396	790,252	67.9	98.7
1996	299,492	701,037	297,832	793,016	68.6	98.8
1997	300,721	679,720	298,479	787,294	67.0	98.8
1998	301,085	680,108	299,053	784,865	67.5	98.8
1999	299,842	785,924	299,842	785,924	88.7	100.0
2000			300,781	785,985		
2001			301,946	785,957		
2002			303,652	787,973		
2003			305,633	791,141		
2004			307,687	794,386		
2005			309,833	797,654		
2006			312,904	803,514		
2007			na	810,222		

Source: Linda (1968P-2007F).

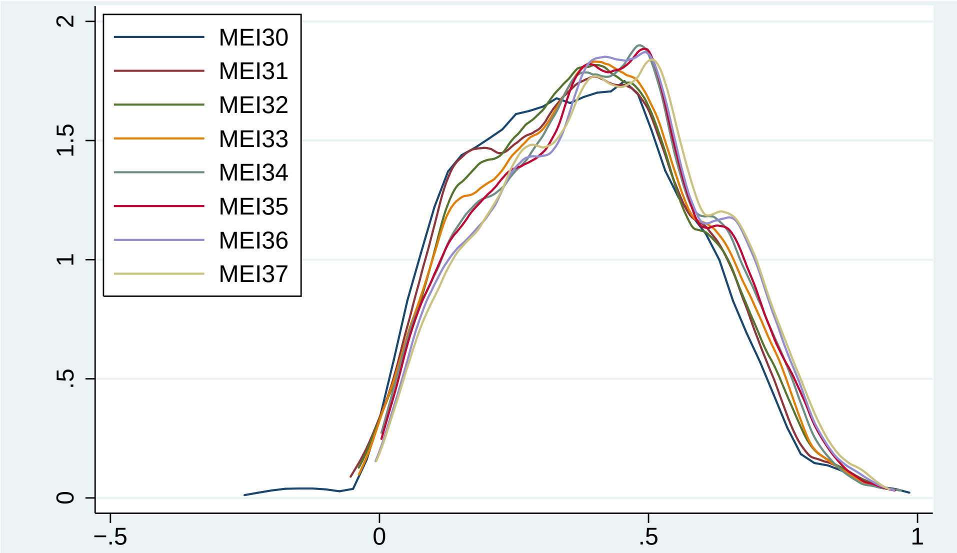

Figure A.1: Distribution of Marginal Effects, 'Top 3%', Varying T (e.g. MEI32–marginal effect, calculated per individual, $T = 32$).

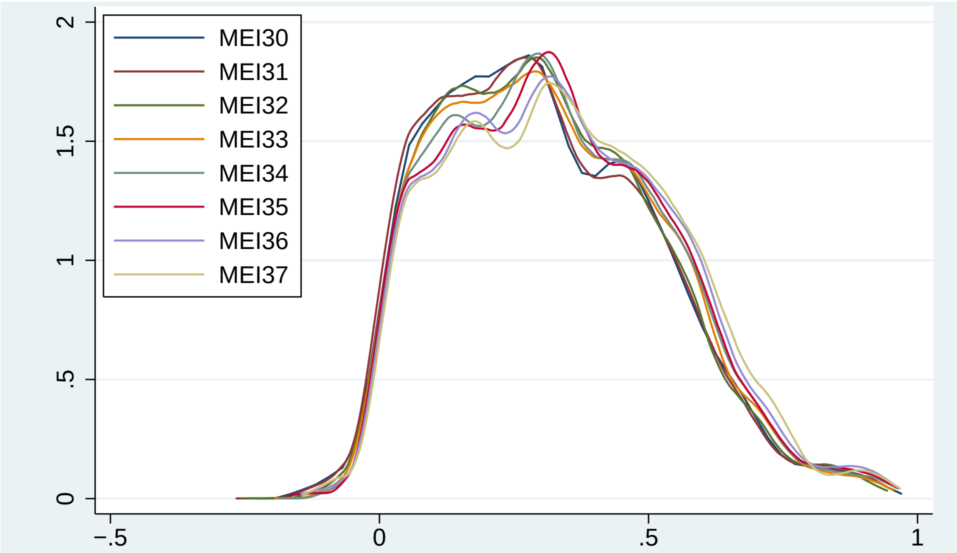

Figure A.2: Distribution of Marginal Effects, 'Millionaires', Varying T (e.g. MEI32–marginal effect, calculated per individual, $T = 32$).

Appendix B The Swedish wealth tax

The Swedish wealth tax in its modern form was introduced in 1948 when there was an extensive tax reform. It has never been a major source of government revenue. The main arguments for it has been equity and redistribution. The wealth tax was repealed from 2007 by the newly elected right-wing government.[21] The tax remains a hot political topic, the Social Democrats say that they will reintroduce the tax if they regain power in the 2010 parliamentary election.

The main features of the wealth tax were unchanged over the main observation period, 1968–2005. It differed from other taxes in that it was levied at the household level and not individual level as the other personal taxes are. This was, in other words, the only example when the household, and not the individual or the firm, was the unit of taxation.[22]

The net wealth of the adult members of the household was added together with the net wealth of the minor children of the household. The tax base was a comprehensive measure of household net wealth (including real assets and financial assets minus debts).[23] Household tax liability was subsequently individualized according to the net wealth share of the individual within the household.

Taxable wealth did not include pension wealth in the sense that the value of future public and occupational pensions were not included neither were savings in tax deferred pension savings accounts. The values of cars, boats, art, and life insurance were not included. In addition there was far from complete coverage of assets abroad.

The wealth tax system was conceptually simple. There was a generous exemption level, exempting, on average, more than 90–95 percent of all households from paying any taxes at all. We refer to this as tax bracket zero with a marginal tax rate of zero. As of 2001, households with two adult spouses got a higher exemption than single households.

Subsequently, (progressively) positive marginal tax rates were applied to subsequent brackets. In later years, the system was been simplified to a two-bracket system with a zero-marginal rate in bracket zero and a single positive one. Tax reforms were discontinuous but frequent and marginal, in that every few years bracket limits have been adjusted, marginal rates have been changed, or the number of tax brackets has been varied. In addition, in all years between nominal changes, the real value of the exemption threshold was affected by inflation (fiscal drag).

Table B.1 reports the main aspects of the Swedish wealth tax exemptions and

[21]The Swedish repeal of the wealth tax followed similar repeals in Austria (from 2001), Denmark (from 1997), Finland (from 2006), and Italy (from 2005).

[22]The personal income tax was joint between spouses before 1971. From 1966 couples could, however, apply to be treated as single filers, see Selin (2009).

[23]The owner to the international clothing retail company H&M threatened to leave the country in 1990s. The government, therefore, introduced some new valuation principles that in practice meant that a handful superrich basically became tax exempt.

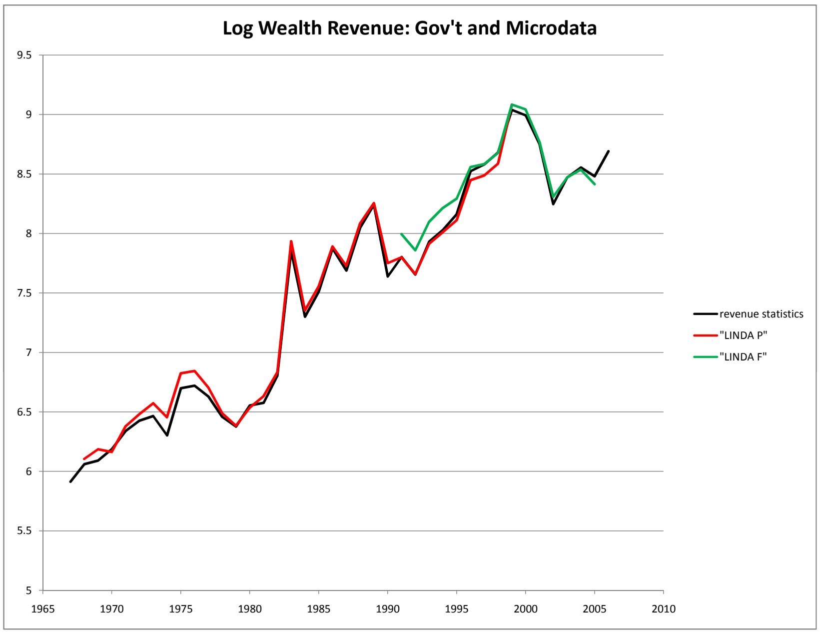

Figure B.1: Wealth tax revenue: Total and micro data, logarithm

rates during the period 1948–2006.

Figure B.1 shows how wealth tax revenue from the households in the micro data set corresponds total wealth tax revenue. It is clear from figure that tax revenue according to the micro data tracks total tax revenue surprisingly well.

Table B.1: Wealth tax rates, 1948–2006.

law	from: Net wealth Dec 31	from: assessment year	exemption, SEK 1,000 singles	exemption, SEK 1,000 couples	n of brackets	top rate, %	source
1947 CGWL	1948	1949	30		5	1.8	SSB 1947:577
	1953	1954	50		6	1.8	SSB 1947:577, SSB 1952:407
	1957	1958	100		6	1.8	SSB 1947:577, SSB 1957:106
	1966	1967	100		5	1.8	SSB 1947:577, SSB 1965:73
	1971	1972	150		4	2.5	SSB 1947:577, SSB 1970:170
	1974	1975	275		4	2.5	SSB 1947:577, SSB 1974:311
	1981	1982	400		4	2.5	SSB 1947:577, SSB 1980:1055
	1983	1984	400		4	2.5	SSB 1947:577, SSB 1982:1190
1983 TCGWL	1983 only	1984 only	300		5	4	SSB 1983:968
	1985	1986	400		4	3	SSB 1947:577, SSB 1984:1080
	1990	1991	800		3	3	SSB 1947:577, SSB 1989:1026
	1991	1992	800		2	2.5	SSB 1947:577, SSB 1991:416
	1992	1993	800		1	1.5	SSB 1947:577, SSB 1992:1489
1997 CGWL	1997	1998	900		1	1.5	SSB 1997:323
	2001	2002	1,000	1,500	1	1.5	SSB 1997:323, SSB 2000:1422
	2002	2003	1,500	2,000	1	1.5	SSB 1997:323, SSB 2001:836
	2005	2006	1,500	3,000	1	1.5	SSB 1997:323, SSB 2004:1039
repeal	2007	2008					SSB 1997:323, SSB 2007:1403

Notes. CGWL is Central Government Wealth Tax Law, TCGWL is Temporary Central Government Wealth Tax Law, SSB (*SFS*) is Swedish Statute Book

Appendix C Estimation

This appendix shows how to estimate a general heterogeneous AR(p) fixed effects probit model by Maximum Likelihood. The main perceived difficulty lies in the fact that it has a very large number of parameters, and that the associated covariance matrix is, therefore, of very large dimensions.

Greene (2004) gives details on estimation by Newton-Raphson making use of the structure of the associated sparse Hessians for a limited dependent variable model with fixed individual constants. Greene refers to Prentice and Gloeckler (1978). We extend the idea to the heterogeneous AR(p) fixed effects probit.

The model is of the form

$$y^{\star}_{it} = \sum_{j=1}^{p} \gamma_i^j y_{i,t-j} + x_{it}\beta + \alpha_i + \varepsilon_{it},$$

where $y^{\star}$ is the latent variable of interest, $y = \mathbf{1}[y^{\star} > 0]$ defines the observed 0/1 indicator, x is a K vector of regressors. Let $i = 1,\ldots,N$ index individuals and $t = 1,\ldots,T$ time. For notational convenience only, we focus on the case of a balanced panel in what follows. Parameters to be estimated are β (a K-vector), α (an N-vector) and p different N-vectors γ^j, $j = 1,\ldots,p$, so that the model has a total of $(p+1)N+K$ parameters. With error ε an NT vector of i.i.d. standard normal variates, the model is a probit.

Define first and second partial derivatives of the log likelihood contribution of observation it as

$$\delta_{it} = \frac{\partial \ln P_{it}}{\partial \beta_0} \quad \text{and} \quad \psi_{it} = \frac{\partial^2 \ln P_{it}}{\partial \beta_0 \partial \beta_0} \tag{12}$$

where β_0 is a (generic) constant and P denotes the probit probability.

Denote the gradient by $g = ({g_\beta}', {g_\alpha}', {g_\gamma^1}', \ldots, {g_\gamma^p}')'$ with $g_\alpha = (g_{\alpha_1}, \ldots, g_{\alpha_N})'$, and $g_{\gamma^j} = (g_{\gamma_1^j}, \ldots, g_{\gamma_N^j})'$, $j = 1,\ldots,p$, and the Hessian as

$$H = \begin{bmatrix} H_{\beta\beta} & H_{\beta\alpha} & H_{\beta\gamma^1} & \cdots & H_{\beta\gamma^p} \\ & H_{\alpha\alpha} & H_{\alpha\gamma^1} & \cdots & H_{\alpha\gamma^p} \\ & & H_{\gamma^1\gamma^1} & \cdots & H_{\gamma^1\gamma^p} \\ & & & \ddots & \vdots \\ & & & & H_{\gamma^p\gamma^p} \end{bmatrix}$$

with blocks such as $H_{\beta\alpha} = ({h_{\beta\alpha_1}}', \ldots, {h_{\beta\alpha_N}}')'$, and so on.

Using (12), elements of the Hessian and the gradient can be found as follows.

$$g_\beta = \sum_i \sum_t \delta_{it} x_{it}, \quad g_{\alpha_i} = \sum_t \delta_{it}, \quad \text{and} \quad g_{\gamma_i^j} = \sum_t \delta_{it} y_{it-j},$$

$$H_{\beta\beta} = \sum_i \sum_t [\psi_{it} x_{it}]' x_{it}, \quad h_{\beta\alpha_i} = \sum_t \psi_{it} x_{it}, \quad h_{\beta\gamma_i^j} = \sum_t \psi_{it} x_{it} y_{it-j},$$

$$h_{\alpha_i\alpha_i} = \sum_t \psi_{it}, \quad h_{\alpha_i\gamma_i^j} = h_{\gamma_i^j\gamma_i^j} = \sum_t \psi_{it} y_{it-j}, \quad \text{and} \quad h_{\gamma_i^j\gamma_i^k} = \sum_t \psi_{it} y_{it-j} y_{it-k}.$$

Also, for $i \neq \iota$,

$$h_{\alpha_i \alpha_\iota} = h_{\alpha_i \gamma_\iota^j} = h_{\gamma_i^j \gamma_\iota^j} = h_{\gamma_i^j \gamma_\iota^k} = 0. \tag{13}$$

Both H and variance-covariance matrix $\Sigma = -H^{-1}$ are of size $((p+1)N+K) \times ((p+1)N+K)$. This matrix can be obtained using inversion of a $K \times K$ matrix and pre- and postmultiplication with a series of matrices of at most order $(p+1)N \times K$, as we will show now.

The implication is that a simple Newton-Raphson algorithm can be used to estimate the model quickly and efficiently. Denote the parameter vector by $\theta = (\beta', \alpha', \gamma^{1'}, \gamma^{2'}, \ldots, \gamma^{p'})'$ then the updating step of parameters between two iterations ℓ and $\ell+1$ is

$$\theta_{\ell+1} = \theta_\ell + s\Sigma g$$

Scalar s is a step size optimization parameter.

Calculation of Σ relies on recursive application of an inversion rule for partitioned matrices, and on the fact that most of the individual blocks of the Hessian are diagonal, owing to independence between units, see (13).

Start with the $N \times N$ lower right submatrix of H, $H_{\gamma^p \gamma^p}$. Its inverse is

$$H^{-1}_{\gamma^p \gamma^p} = \begin{bmatrix} 1/h_{\gamma_1^p \gamma_1^p} & 0 & \cdots \\ 0 & \ddots & \vdots \\ \vdots & 0 & 1/h_{\gamma_i^p \gamma_i^p} \end{bmatrix}$$

and $h_{\gamma_i^p \gamma_i^p}$ denotes the ith diagonal element of $H_{\gamma^p \gamma^p}$. Consider next the coefficients γ^q where $q = p-1$. Then the inverse of the $2N \times 2N$ matrix

$$\left[\begin{array}{c|c} H_{\gamma^q \gamma^q} & H_{\gamma^q \gamma^p} \\ \hline H_{\gamma^q \gamma^p} & H_{\gamma^p \gamma^p} \end{array}\right]$$

will be denoted

$$\left[\begin{array}{c|c} H^{\gamma^q \gamma^q} & H^{\gamma^q \gamma^p} \\ \hline H^{\gamma^q \gamma^p} & H^{\gamma^p \gamma^p} \end{array}\right]$$

and the various blocks can be computed as

$$\begin{array}{rcl} H^{\gamma^q \gamma^q} & = & (H_{\gamma^q \gamma^q} - H_{\gamma^q \gamma^p} H^{-1}_{\gamma^p \gamma^p} H_{\gamma^q \gamma^p})^{-1} \\ H^{\gamma^q \gamma^p} & = & -H^{-1}_{\gamma^p \gamma^p} H_{\gamma^q \gamma^p} H^{\gamma^q \gamma^q} \\ H^{\gamma^p \gamma^p} & = & H^{-1}_{\gamma^p \gamma^p}(I_N + H_{\gamma^q \gamma^p} H^{\gamma^q \gamma^p}). \end{array} \tag{14}$$

There is no numerical inversion involved as all matrices (including $H^{\gamma^q \gamma^q}$) are diagonal. Multiplication of these diagonal matrices can be performed on vectors.

One level up, denote $r = p-2$ and find the inverse of

$$\left[\begin{array}{c|cc} H_{\gamma^r \gamma^r} & H_{\gamma^r \gamma^q} & H_{\gamma^r \gamma^p} \\ \hline \cdot & H_{\gamma^q \gamma^q} & H_{\gamma^q \gamma^p} \\ \cdot & \cdot & H_{\gamma^p \gamma^p} \end{array}\right]$$

It is immediately obvious that we can again calculate the inverse without numerical inversion, applying the rule used in (14) again. Thus we can proceed until lag $p-(p-1)$. Let the Hessian at that stage be denoted by $H_{\gamma\gamma}$. The next item in the sequence then will involve

$$H_{\alpha\gamma\alpha} = \left[\begin{array}{c|ccc} H_{\alpha\alpha} & H_{\alpha\gamma^1} & \cdots & H_{\alpha\gamma^p} \\ \hline H_{\alpha\gamma^1} & H_{\gamma^1\gamma^1} & \cdots & H_{\gamma^1\gamma^p} \\ \vdots & \vdots & \ddots & \vdots \\ H_{\alpha\gamma^p} & & \cdots & H_{\gamma^p\gamma^p} \end{array}\right]$$

The inverse of $H_{\alpha\gamma\alpha}$ can be written as

$$H_{\alpha\gamma\alpha}^{-1} = \left[\begin{array}{c|c} H^{\alpha\alpha} & H^{\alpha\gamma} \\ \hline H^{\alpha\gamma'} & H^{\gamma\gamma} \end{array}\right]$$

where

$$\begin{aligned} H^{\alpha\alpha} &= (H_{\alpha\alpha} - H_{\alpha\gamma}H_{\gamma\gamma}^{-1}H'_{\alpha\gamma})^{-1} \\ H^{\alpha\gamma} &= [-H_{\gamma\gamma}^{-1}H'_{\alpha\gamma}H^{\alpha\alpha}]' \\ H^{\gamma\gamma} &= H_{\gamma\gamma}^{-1}(I_{pN} - H'_{\alpha\gamma}H^{\alpha\gamma}) \end{aligned}$$

and $H_{\alpha\gamma}$ is the $N \times pN$ upper right submatrix of $H_{\alpha\gamma\alpha}$; $H_{\alpha\alpha}$ is $N \times N$ diagonal, hence, $H_{\alpha\gamma\alpha}^{-1}$ can be obtained without numerical inversion, too.

Finally, the complete Hessian shall be written

$$H = \left[\begin{array}{c|c} H_{\beta\beta} & H_{\beta\alpha\gamma} \\ \hline H_{\beta\alpha\gamma}{}' & H_{\alpha\gamma\alpha} \end{array}\right]$$

with inverse

$$H^{-1} = \left[\begin{array}{c|c} H^{\beta\beta} & H^{\beta\alpha\gamma} \\ \hline H^{\beta\alpha\gamma'} & H^{\alpha\gamma\alpha} \end{array}\right]$$

where $H^{\beta\alpha\gamma}$ and $H_{\beta\alpha\gamma}$ are the upper right $K \times (p+1)N$ submatrices of H^{-1} and H, respectively. This inverse, then, consists of blocks

$$\begin{aligned} H^{\beta\beta} &= \left[H_{\beta\beta} - H_{\beta\alpha\gamma}H_{\alpha\gamma\alpha}^{-1}H'_{\beta\alpha\gamma}\right]^{-1} \\ H^{\beta\alpha\gamma} &= [-H_{\alpha\gamma\alpha}^{-1}H'_{\beta\alpha\gamma}H^{\beta\beta}]' \\ H^{\alpha\gamma\alpha} &= H_{\alpha\gamma\alpha}^{-1}\left[I_{(p+1)N} - H'_{\beta\alpha\gamma}H^{\beta\alpha\gamma}\right] \end{aligned}$$

As $H_{\beta\beta}$ is not diagonal, calculating $H^{\beta\beta}$ involves numerical inversion of a $K \times K$ matrix.