The Role of Wealth Transformations: An Application to Estimating the Effect of Tax Incentives on Saving

Karen M. Pence*
The Role of Wealth Transformations: An Application to Estimating the Effect of Tax Incentives on Saving∗

Karen M. Pence

Abstract

Researchers may want to estimate the percentage change of a variable, such as household wealth or corporate profits, that takes on economically significant nonpositive values. Using the logarithmic transformation, however, requires discarding observations with nonpositive values. This paper describes a possible solution to this problem—the inverse hyperbolic sine transformation—and shows how to implement this transformation optimally in the case of median regression. As an illustration of the usefulness of this transformation, I revisit a specification sometimes used to estimate the effect of tax incentives on household saving.

KEYWORDS: inverse hyperbolic sine, transformations, 401(k) program

∗Federal Reserve Board of Governors MS 93 Washington, DC 20551 karen.pence@frb.gov I am grateful to John Karl Scholz, Yuichi Kitamura, and Bob Haveman for their guidance and to the editor, two referees, Bill Gale, John Sabelhaus, and numerous colleagues at the Federal Reserve Board and elsewhere for helpful comments. I thank especially the Survey of Consumer Finances staff for their work developing and documenting the survey and Carolyn Aler, Mary DiCarlantonio, and Michele Berrios for outstanding research assistance. The Christensen Award in Empirical Economics and the Social Science Research Council Program in Applied Economics, funded by the MacArthur Foundation, provided generous financial support. Part of this paper circulated earlier under the title “401(k)s and Household Saving: New Evidence from the Survey of Consumer Finances.” The views presented are not necessarily those of the Federal Reserve Board, the Social Science Research Council, or the MacArthur Foundation.
INTRODUCTION

Several variables commonly used in economics, including household wealth, corporate profits, and net exports, can take on economically significant nonpositive values. This paper highlights the usefulness of the inverse hyperbolic sine (IHS) transformation for estimating percentage change specifications with these variables. In particular, I use the transformation to revisit a specification sometimes used to estimate the effect of tax incentives such as individual retirement accounts and 401(k) accounts on household saving.

The specification compares the change in the level of wealth over time of households eligible for the tax incentive to that of ineligible households. However, even after controlling for observed characteristics, eligible households may start the time period with more wealth than ineligible households. As a result, the wealth accumulation of the two groups can differ both because they exhibit different saving behavior and because eligible households, due to their higher initial wealth, have greater investment gains than ineligible households.

The inverse hyperbolic sine transformation can help address this “difference in initial wealth” problem. As is shown in the paper, when the two groups differ in their initial wealth and when investment returns are a sizeable part of the change in wealth, a proportionate specification may isolate better the effect of tax incentives on wealth accumulation. Otherwise, a levels specification may be preferable. Depending on the chosen value of a scaling parameter, the IHS can approximate either the level or logarithm of wealth. Unlike the logarithm, however, the IHS is defined for households with nonpositive wealth. I show how to estimate the scaling parameter that provides the best fit to the data.

As an empirical example, I use the 1995, 1998, and 2001 Surveys of Consumer Finances to estimate the effect of 401(k) eligibility on two wealth measures: net financial assets and net financial assets excluding 401(k) balances. These two measures showcase the conditions under which the IHS transformation can provide different estimates than the levels transformation.

In the case of net financial assets, eligible households held considerably more of these assets than ineligible households in 1995 and realized substantial capital gains on these assets over the 1995-2001 period. As a result, the levels estimates are considerably larger than the IHS estimates. The levels estimates suggest that the net financial assets of eligible households increased by $6,690 relative to ineligible households over the 1995–2001 period, whereas the IHS estimates indicate that the relative increase was only $1,130.

In the case of net financial assets excluding 401(k) balances, eligible and ineligible households had about the same amount of these assets in 1995. In addition, capital gains on these assets were relatively modest over the 1995-2001 period because most households hold the bulk of their stock within their 401(k)
accounts. In this situation, the levels and IHS estimates provide near-identical answers. Both specifications indicate that eligible and ineligible households accumulated about the same amount of net financial assets, excluding 401(k) balances, over this period.

In the case of falling asset values, the IHS estimates may be larger than the levels estimates. For example, since eligible households have more housing wealth than ineligible households, a levels specification estimated over the early 1990s, when real housing values fell, might show a larger relative decrease in the net worth of eligible households than an IHS specification. In other applications, such as corporate profits, the IHS may likewise be a useful transformation in times of both rising and falling profitability.

**ESTIMATING THE EFFECT OF TAX INCENTIVES ON SAVING**

Researchers commonly assess the effect of tax incentives on saving by comparing the wealth of cohorts of eligible and ineligible households. These cohort comparisons generally use a difference-in-difference framework in which \( w \) is a measure of the level of wealth, \( \text{eligible} \) is a dummy variable indicating eligibility, \( \text{year}_2 \) is a dummy variable indicating that the household was surveyed in the latter of two survey years, and \( X \) contains other covariates that are correlated with wealth accumulation

\[
w = \phi \cdot \text{eligible} + \delta \cdot \text{year}_2 + \lambda \cdot \text{eligible} \cdot \text{year}_2 + X \beta + \varepsilon.
\]

If the program increases saving, the wealth of eligible households should increase more over time than the wealth of ineligible households, or \( \lambda > 0 \). The underlying identification assumption is that without the tax incentives, eligible and ineligible households would have equivalent wealth accumulation patterns after controlling for the explanatory variables.\(^1\)

This difference-in-difference approach is popular because appropriate panel data are often unavailable. Variants of this specification have been used by Venti and Wise (1992, 1995), Engen, Gale, and Scholz (1994), and Attanasio and DeLeire (2002) to assess the effectiveness of IRAs; by Poterba, Venti, and Wise (1995, 1996, 1998), Engen, Gale, and Scholz (1994, 1996), Engen and Gale (1997, 2000), Sabelhaus and Ayotte (1997), and Engelhardt (2001) to assess the effectiveness of 401(k)s; by Engelhardt (1996) to assess the effectiveness of

\(^1\)As is widely noted in this literature, this identification assumption is violated if eligible households have a greater taste for saving than ineligible households. I do not address this issue in this paper. Engen and Gale (2000) provide an overview of this issue.
Canada’s Registered Home Ownership Savings Plan; and by Ma (2004) to assess the effectiveness of Coverdell education saving accounts and 529 plans.2

Because changes in the level of wealth include both asset returns and contributions, whereas saving is generally defined only as contributions, authors in this literature have disagreed on how to interpret changes in wealth when one group starts the time period with more wealth. For example, Poterba, Venti, and Wise (1995) attribute the increase in the net financial assets of 401(k)-eligible households relative to ineligible households over the 1987–91 period to the saving incentives of the 401(k) program; Engen and Gale (2000) attribute it to the greater financial holdings of eligible households and the rise in the stock market. Engen and Gale (1997) interpret the drop in the net worth of 401(k)-eligible households, relative to ineligible households, over the 1987–91 period as evidence that 401(k) accounts do not increase saving; Poterba, Venti, and Wise (1998) interpret it as evidence that eligible households had more housing wealth than ineligible households and so were affected disproportionately by the drop in housing values over this period.

To see these concerns more formally, note that the difference-in-difference coefficient \( \lambda \), abstracting from the other covariates, can be written as follows, where the superscripts \{e, i\} represent eligible and ineligible households:

\[
\lambda = (w_t^e - w_{t-1}^e) - (w_t^i - w_{t-1}^i).
\]

If \( r \) is the return on existing assets and \( c_t \) is contributions, then \( w_t = (1 + r)w_{t-1} + c_t \). Assuming eligible and ineligible households earn the same rate of return on their investments,

\[
\lambda = (c_t^e - c_t^i) + r(w_{t-1}^e - w_{t-1}^i)
\]

and the coefficient includes both differences in the contributions of the two groups and differences in their investment earnings.3 In contrast, in a specification in which the dependent variable is transformed with the logarithm,

\[
\hat{\lambda} \approx \frac{w_t^e - w_{t-1}^e}{w_t^e - w_{t-1}^e} \cdot \frac{w_t^i - w_{t-1}^i}{w_t^i - w_{t-1}^i} = \frac{c_t^e}{w_t^e - w_{t-1}^e} \cdot \frac{c_t^i}{w_t^i - w_{t-1}^i}
\]

---

2 See Bernheim (2002) for a review of the taxation and saving literature. Studies of the effect of tax incentives on saving that do not use a difference-in-difference framework include Venti and Wise (1990), Gale and Scholz (1994), Benjamin (2003), and Chernozhukov and Hansen (2004).

3 The rate of return received by eligible households, \( r^e \), may be higher than that received by ineligible households, \( r^i \), because eligible households receive favorable tax treatment on their tax-favored accounts. In this case, \( \lambda \) will also include the term \((r^e w_{t-1}^e - r^e w_{t-1}^i) - (r^i w_{t-1}^i - w_{t-1}^i)\) in the levels specification and the term \((r^e - r^i)\) in the log specification.
and \( \lambda \) reflects differences in contributions (scaled by each group’s wealth in the previous period) but not differences in investment earnings.

When eligible and ineligible households start the time period with the same amount of wealth, both the levels and logs regressions will indicate the effect of saving incentives on contributions. In a multivariate regression, this condition is met if the wealth of eligible and ineligible households, conditional on the explanatory variables, is comparable. Otherwise, which specification is preferable depends on whether most of the change in wealth comes from contributions or from asset returns. If the change in wealth comes entirely from contributions, the levels specification may be better if the researcher does not believe that contributions are proportional to prior-period wealth. If the change in wealth comes from asset returns, the log specification may be better because it controls for the extra returns earned by the group with higher prior-period wealth.

However, in order to implement a log-wealth specification, researchers must drop from the sample households with more liabilities than assets or no financial wealth at all. These households can make up a sizeable fraction of the sample: twenty-one percent of households in the 2001 Survey of Consumer Finances, for example, have net financial assets valued at or below zero. The inverse hyperbolic sine transformation, described in the next section, provides a way to estimate a percentage change specification without excluding negative-wealth households.

Although this paper is the first to propose the inverse hyperbolic sine as a solution to the difference-in-initial-wealth problem, other researchers have recognized this issue and taken steps to address it. Poterba, Venti, and Wise (1995), for example, establish that in 1984—the starting year in their analysis—eligible and ineligible households had about the same amount of financial assets, conditional on income. Engen and Gale (2000) estimate log wealth specifications for groups whose median wealth exceeds zero, after first replacing negative and zero wealth values with the value “1.” Other studies focus on variables such as contributions (Venti and Wise, 1995) or consumption (Attanasio and DeLeire, 2002) whose initial values are likely to be more comparable across groups than wealth.

In fact, if specifications from previous studies were re-estimated with the inverse hyperbolic sine, only a few might see their results change. Engen, Gale, and Scholz (1994) and Sabelhaus and Ayotte (1998), for example, run levels specifications on groups that may have unequal initial wealth. (These studies also include other specifications not affected by the difference-in-initial wealth problem.) However, even studies that control adequately for differences in initial wealth might benefit from the IHS. For example, researchers who addressed this
issue by limiting their samples to groups with comparable initial wealth now have
the flexibility to also compare groups for which this condition does not hold.

THE INVERSE HYPERBOLIC SINE AND MEDIAN REGRESSION

The inverse hyperbolic sine. The inverse hyperbolic sine (IHS) of wealth takes
the following form, in which \( \theta \) is a scaling parameter and \( w \) is a measure of
wealth: 4

\[
\theta^{-1} \sinh^{-1}(\theta w) = \theta^{-1} \ln(\theta w + (\theta^2 w^2 + 1)^{1/2}).
\]

This symmetric function is linear around the origin and approximates the
logarithm in its right tail. In fact, for large \( w \), this function is simply a vertical
displacement of the logarithm: \( \ln(\theta w + (\theta^2 w^2 + 1)^{1/2}) \approx \ln2 + \ln w \).

The scaling parameter \( \theta \) governs the proportion of the function’s domain
in which it is essentially linear and the proportion in which it approximates the
logarithm. As \( \theta \) approaches zero, the IHS transformation is linear for a larger
proportion of its domain. As shown in Figure 1, the IHS transformation is linear
for less of its domain when it is evaluated at \( \theta = 0.001 \) than at \( \theta = 0.0003 \) (the
optimal value in this paper, as estimated by maximum likelihood).

To see this relationship another way, note that the derivative of \( \text{IHS}(w) \) is
\[ 1/\sqrt{\theta^2 w^2 + 1} \]. If \( w \) is large relative to \( \theta \), this derivative approximates the
derivative of the log, \( 1/w \), for most positive values of \( w \). If \( w \) is small relative
to \( \theta \), the derivative is approximately one and the function is linear for most of its
domain.

The linearity of the IHS through the origin may be an appealing
assumption for small values of wealth: we may not like the logarithm’s feature of
treating a household with a wealth increase from $1 to $2 equivalently to a
household with a wealth increase from $10,000 to $20,000. In addition, the shape
of the IHS may approximate well the relationship between wealth, contributions,
and investment returns. Specifically, for households with small amounts of
wealth, changes in wealth may be more likely to stem from contributions, and be
described best by a linear relationship. For households with large amounts of
wealth, increases in wealth may be more likely to stem from investment returns,
and be described best by a logarithmic relationship.

---

4 See Burbidge, Magee, and Robb (1988) for a more detailed discussion of the inverse hyperbolic
sine. Johnson (1949) originally proposed this transformation.

Contributions to Economic Analysis & Policy
The coefficients from the model \( y = \text{IHS}(w) = X\beta + \varepsilon \) can be expressed as the marginal effects of a change in a given variable \( x \) on dollars of wealth \( w \) by applying the transformation \( \frac{1}{2}(e^{\theta y} + e^{-\theta y})\beta \). This transformation is based on the relationship \( \frac{dw}{dx} = \frac{dw}{dy} \frac{dy}{dx} \frac{1}{\beta} \). By definition, \( w \) is equivalent to the hyperbolic sine, \( \sinh \), of \( y \), since \( w = \sinh(\sinh^{-1}(w)) = \sinh(y) \):

\[
 w = \frac{1}{2\theta} [e^{\theta y} - e^{-\theta y}] \quad \Rightarrow \quad \frac{dw}{dy} = \frac{1}{2} [e^{\theta y} + e^{-\theta y}].
\]

Alternatively, when \( \beta \) is multiplied by \( \theta \), it approximates (for large \( w \)) the effect of a unit change in \( x \) on the percentage change in \( w \).\(^5\) This interpretation is akin to that in a regression when the dependent variable is transformed with the log. To see this relationship, note that if \( \theta w \) is sufficiently large,

\(^5\) This approximation only works for the part of the function’s domain in which it resembles the logarithm.
Another transformation that is defined for positive and negative values, and nests both the linear and logarithmic transformation, is the Box-Cox transformation. For wealth data, however, the IHS is preferable because the Box-Cox transformation is not defined at zero. In my sample of SCF respondents, five percent have zero net financial assets. Burbidge, Magee and Robb (1988) describe the merits of both transformations and note that the IHS provided a better fit than Box-Cox to their sample of Canadian net worth data.

**Median regression.** Median regression is commonly used in studies of the effectiveness of tax incentives because it is less sensitive than ordinary least squares to the extreme values that are characteristic of wealth data. Median regression minimizes the sum of absolute deviations and its coefficients express the marginal effect of one variable on the conditional median of another. Extreme values affect median regression less than mean regression because median regression coefficients are determined by the order of the data points rather than their magnitude.

Another advantage of median regression is that, unlike ordinary least squares (OLS), the difference-in-difference coefficient is an unbiased estimator of percentage change. In the OLS framework, the semi-elasticity of the regression function \( E(w|x) \) with respect to a one-unit change in \( x \) does not equal the quantity typically estimated by researchers. More formally, \(^6\)

\[
\frac{\partial E(w|x)}{\partial x_i} = \frac{1}{E(w|x)} \frac{\partial \log E(w|x)}{\partial x_i} \neq \frac{\partial \log E(\log(w)|x)}{\partial x_i}.
\]

However, the latter equality holds for median regression because of the property that for any monotonic function \( f \), \( \text{median}(f(x)) = f(\text{median}(x)) \), or equivalently,

\[
\frac{\partial \log(\text{med}(w|x))}{\partial x_i} = \frac{\partial \text{med}(\log(w)|x)}{\partial x_i}.
\]

This result holds as well for the IHS transformation because it is also monotonic.

Median regression has two disadvantages — heteroskedasticity and nonconvergence — that both can be ameliorated with the IHS transformation. In the presence of heteroskedasticity, the textbook standard error formula for median regression can understate the true standard errors (Rogers (1992)).

\(^6\) See Wooldridge (2002) p. 17 for more discussion.
Heteroskedasticity can be an issue for wealth data regressions because residuals are often larger for high-income households than low-income households. Nonconvergence can be an issue because the algorithms that estimate median regression can have numerical overflow problems when the data contain outliers. In this situation, the estimates may either fail to converge or take on implausible values. By damping outliers, the IHS transformation addresses both of these problems.\(^7\)

To estimate the optimal \(\theta\), I exploit the fact that estimating the model \(\text{IHS}(w, \theta) = X\beta + \varepsilon\) with median regression is equivalent to estimating it with maximum likelihood if \(\varepsilon\) has a Laplace, or double exponential, distribution. The log-likelihood takes the following form, in which \(n\) is the sample size: \(^8\)

\[
L(\theta, \beta) = -n \ln \left( \frac{2}{n} \right) - n \ln \left( \sum_{i=1}^{n} |\text{IHS}(w, \theta) - X\beta| \right) - n - \sum_{i=1}^{n} \ln \left( \sqrt{\theta^2 w^2 + 1} \right).
\]

A similar likelihood function can be derived for OLS if \(\varepsilon\) has a normal distribution (Burbidge, Magee, and Robb, 1989). As \(\theta \to 0\), \(\text{IHS}(w) \to w\), so it is possible to see whether the levels of wealth specification provides the best fit to the data by testing whether the optimal \(\theta\) is statistically different from zero.

As noted earlier, Engen and Gale (2000) propose implementing a percentage change specification by replacing wealth data points that take on negative and zero values with the value 1 and then applying the logarithm transformation.\(^9\) Essentially, this method assumes that the conditional median of the original wealth distribution is the same as the conditional median of the same wealth distribution with its left tail replaced with a mass point at one. For the high-income groups that Engen and Gale (2000) examine, this assumption is likely to be accurate. However, because this method reduces the variation in the data, it will understate the true median regression standard errors and does not

---

\(^7\) Because median regression is scale-invariant, researchers may also be able to obtain convergence – and retain the original point estimates – by dividing the wealth measure by a large constant.

\(^8\) To incorporate sample weights, I estimate a weighted log-likelihood in which \(q\) is the original sample weight and \(q^*\) is the normalized sample weight \(nq/\Sigma q\).

\[
L(\theta, \beta) = -n \ln \left( \frac{2}{n} \right) - n \ln \left( \sum_{i=1}^{n} q^* |\text{IHS}(w, \theta) - X\beta| \right) - n - \sum_{i=1}^{n} q^* \ln \left( \sqrt{\theta^2 w^2 + 1} \right).
\]

Weights are also used to estimate the median regression residuals \(|\text{IHS}(w, \theta) - X\beta|\).

\(^9\) In their studies of the black-white wage gap, Johnson, Kitamura, and Neal (2000) and Chandra (2003) use a similar method to impute the shadow log wages of individuals who are not working.
allow researchers to identify the response of low-income households to saving incentives.\textsuperscript{10}

The inverse hyperbolic sine transformation solves the problem of negative wealth values without restricting the sample or distorting the standard errors. Although a few authors have used the IHS in other applications, this study is the first to show in detail its usefulness for implementing a percentage change specification and to demonstrate how to implement the IHS optimally within a median regression framework.\textsuperscript{11}

**USING THE INVERSE HYPERBOLIC SINE TO ESTIMATE THE EFFECT OF THE 401(K) PROGRAM ON SAVING**

I estimate the effect of the 401(k) program on saving over the 1995–2001 period with the following specification in which $w$ is a measure of wealth, $year_{98}$, $year_{01}$, and eligible are dummy variables, $X$ contains other explanatory variables, and the median of $\varepsilon$, conditional on the explanatory variables, is zero:

\[
w = \phi \cdot \text{eligible} + \lambda_{98} \cdot \text{eligible} \cdot year_{98} + \lambda_{01} \cdot \text{eligible} \cdot year_{01} + \delta_{98} \cdot year_{98} + \delta_{01} \cdot year_{01} + X \beta + \varepsilon.
\]

This framework provides two tests of the hypothesis that 401(k) eligibility increases saving. First, the wealth of eligible households should increase more over time than the wealth of ineligible households. In this case, $\lambda_{98}$ and $\lambda_{01}$ should be positive and $\lambda_{01} > \lambda_{98}$. Second, 401(k) balances should not completely displace other assets. To test this proposition, I examine the change in “wealth excluding 401(k) balances” for eligible households relative to ineligible households. If 401(k)s do not displace other financial assets, $\lambda_{98}$ and $\lambda_{01}$ should be near zero or even positive; if 401(k)s displace some or all of these assets, $\lambda_{98}$ and $\lambda_{01}$ should be negative.

The specification is estimated using data from the Survey of Consumer Finances. The Survey of Consumer Finances (SCF) measures household wealth and financial decisions and contains extensive information on all aspects of the household balance sheet. The Federal Reserve Board has conducted the survey

\textsuperscript{10}Truncating the data decreases the standard errors by reducing the variation in both $w$ and $\varepsilon$. In the specific case of median regression, standard errors are inversely proportional to the density of the errors at zero. Truncating the data may increase this density and thus decrease the standard errors.

\textsuperscript{11}See Burbidge, Magee, and Robb (1988), Carroll, Dynan, and Krane (2003), Kennickell and Sundén (1997), and Kapteyn and Panis (2003) for other studies that use the IHS.

*Contributions to Economic Analysis & Policy*
every three years since 1983; the design of the survey has been consistent since 1989.\textsuperscript{12}

Most previous research in the 401(k) literature has used the Survey of Income and Program Participation (SIPP).\textsuperscript{13} The SIPP has a larger sample size than the SCF and, as discussed below, had a better definition of 401(k) eligibility than the SCF for the waves before 1995. However, the SCF is generally considered the highest-quality wealth data available and also provides tools for addressing three technical issues that can substantively affect standard errors in these regressions: heteroskedasticity, imputation uncertainty, and survey design.\textsuperscript{14}

\textbf{Wealth measures.} I examine the effect of 401(k) eligibility on net financial assets, defined as the sum of savings, checking, and money market accounts, certificates of deposit, stocks, bonds, mutual funds, 401(k)s, other liquid defined contribution plan balances, IRAs, and the cash value of whole life insurance policies, minus credit card balances and other consumer loans. Tangible assets such as real estate and cars, as well as loans against these assets, are not included in this wealth measure. Net financial assets are inflated to 2001 dollars in all years with the current methods consumer price index.

If borrowers finance their 401(k) contributions by decreasing their saving in tangible assets such as home equity, the net financial assets measure could overstate the effect of the 401(k) program on total wealth accumulation. Nonetheless, I focus on net financial assets for three reasons. First, because almost all papers in this literature present at least one net financial assets specification, the results here are easily compared to other studies. Second, because the eligible-ineligible gap in net financial assets is larger in percentage terms than the gap in net worth, the net financial assets measure better demonstrates the strengths of the inverse hyperbolic sine. Third, because financial assets are easier to value than tangible assets, net financial assets is measured more precisely than net worth. Standard errors from net worth specifications are often large enough that no hypothesis can be rejected.

\textbf{401(k) Eligibility.} Unlike the SIPP, the SCF does not ask households directly if they are eligible for the 401(k) program. An accurate measure of 401(k) eligibility can be constructed from the 1995, 1998, and 2001 surveys, but the measure based on the 1989 and 1992 surveys misclassifies approximately

\textsuperscript{12} Aizcorbe, Kennickell, and Moore (2003) provide an overview of the 2001 SCF.
\textsuperscript{14} Curtin, Juster, and Morgan (1989) compare the 1983 SCF to the 1984 Survey of Income and Program Dynamics and the 1984 Panel Study of Income Dynamics and conclude that the SCF is the best overall source of wealth data. Czajka, Jacobson, and Cody (2004) document that the wealth data in the 1996 SIPP panel appears to be of poorer quality than both the SCF and earlier SIPP panels.
eight percent of households. Specifically, if a worker participates in one pension, but decides not to participate in the 401(k) offered by his employer, he is classified as 401(k) ineligible in the 1989 and 1992 surveys. As a result, the empirical work in this paper draws on the 1995–2001 surveys.\(^{15}\)

**Other explanatory variables.** I control for characteristics of workers and their employers that may be correlated with wealth accumulation. Specifically, I control for the age, education, marital status, and race of the household head as well as for household income, the presence of two earners, and defined benefit pension coverage.\(^{16}\) Since large firms and firms with predominantly white-collar workers are more likely to offer a 401(k) program to their workers, I also add variables for the industry, occupation, firm size, and unionization status of the job held by the primary worker in the household.\(^{17}\)

**Sample construction.** The sample consists of households in the pooled 1995, 1998, and 2001 SCF cross-sections in which the head was between the ages of 21 and 58 in 1995 (and hence between the ages of 27 and 64 in 2001). I exclude households with a head older than 64 because their saving decisions might be affected by their option to retire. I keep only households in which at least one spouse is working and neither spouse is self-employed. Since the 401(k) program is a work-based program, individuals who are not in the labor force cannot be eligible for the 401(k) program. Self-employed workers are not eligible for 401(k) plans and often save in a different manner than other workers as well.

The percent of the sample eligible for a 401(k) increased from 54 percent in 1995 to 62 percent in 2001. Although some 401(k)-eligible households in 2001 were not eligible in earlier years, each cross-section has more exposure to the 401(k) program than its predecessor. For example, the median number of years a worker had been participating in a 401(k) at his current job increased from 2 years in 1995 to 4 years in 2001, whereas mean years of participation rose from 4 years to 6 years. In addition, the characteristics of eligible and ineligible households were generally stable across time, with the exception of defined benefit pension coverage, which fell steadily for both groups across the 1990s.

**Sample weights.** I weight the specification because wealthy households are overrepresented in the survey. When the dependent variable (here, wealth) is correlated with the sampling scheme, the median regression estimator is not consistent because the identifying assumption \(\text{med}(\epsilon|X) = 0\) does not hold.\(^{18}\)

---

\(^{15}\) See Pence (2002) for more details on defining 401(k) eligibility in the Survey of Consumer Finances.

\(^{16}\) Household income includes wages, business income, investment income, capital gains, welfare, child support, rental income, and retirement payments.

\(^{17}\) The primary worker is the household head if he or she works and is the head’s spouse otherwise.

\(^{18}\) Holt et al. (1980) and Nathan and Holt (1980) make this argument in more detail for ordinary least squares.
However, if the sampling weights represent the underlying population probabilities accurately, a weighted median regression estimator will be consistent.

**Adjusting the standard errors for heteroskedasticity, imputation uncertainty, and survey design.** Heteroskedasticity, imputation uncertainty, and survey design issues can affect standard error estimates. As mentioned earlier, in a levels-of-wealth regression, residuals are much larger for high-income households than for low-income households; the analytic formula for median regression standard errors, in the presence of this heteroskedasticity, can yield inaccurate estimates. Imputation uncertainty is an issue because many households are reluctant to reveal their holdings of certain assets or simply do not know their values. In these cases, the missing data values must be imputed, and the standard errors should reflect the uncertainty inherent in the imputation. Finally, the standard errors should also take into account the stratification and clustering in the SCF sample design.

Because the SIPP does not provide the appropriate tools, previous papers in this literature have not adjusted their estimates for either imputation uncertainty or the survey design. In addition, the standard errors may be inaccurate in Poterba, Venti, and Wise (1995, 1996, 1998) because they are based on the analytic formula. The SCF, however, provides appropriate tools in the form of 999 bootstrap replicates drawn in accordance with the sample design, along with a vector of sampling weights for each replicate. I use these replicates to bootstrap the standard errors and thus incorporate both heteroskedasticity and the survey design into the estimates. I also use the repeated imputation inference tools provided by the SCF to adjust the standard errors and regression coefficients for imputation uncertainty.

**Results**

Table 1 shows coefficient estimates for the key parameters of interest—*eligible*, *eligible*\text{•}year\textsubscript{98}, and *eligible*\text{•}year\textsubscript{01}—for both the net financial assets and the net financial assets excluding 401(k) balances specifications. Levels and inverse

---

19 For example, only 64 percent of stockowners in the 1995 SCF provided the dollar value of their stock (Kennickell, 1997).

20 See Kennickell (2000) for more information on variance estimation procedures with the SCF. Because of concerns about respondent confidentiality, the SCF does not release data on the stratum and cluster associated with each observation. Researchers may want to account for these survey design features when estimating standard errors. The bootstrap replicates provided by the SCF offer a way to account for the survey design while still protecting respondent confidentiality.

hyperbolic sine coefficient estimates are shown for both wealth measures. The full set of coefficients for the levels and inverse hyperbolic sine transformations are shown in Table 2 for the net financial assets specification only.

**Net financial assets.** Because two key conditions hold, I expect estimates of the effect of 401(k) eligibility on the accumulation of net financial assets over the 1995–2001 period to be larger in the levels specification than the IHS specification. First, eligible households had more net financial assets than ineligible households in 1995: median net financial assets were $22,493 for eligible households and $1,153 for comparable ineligible households. Second, because of the substantial run-up in the stock market, most of the change in wealth over the 1995-2001 period came from investment returns.

As expected, in the levels net financial assets specification, 401(k) eligibility is associated with large increases in wealth. Holding demographic characteristics constant, eligible households held $4,523 more in net financial assets than ineligible households in 1995 (eligible coefficient, Table 1). Over the 1995–98 period, the net financial assets of eligible households, relative to ineligible households, increased $3,266 more in real dollars (eligible•year_{98} coefficient, Table 1); over the 1995–01 period, their net financial assets increased, relative to ineligible households, by $6,690 (eligible•year_{01} coefficient, Table 1). These coefficients are all statistically significant, albeit at a marginal level for eligible•year_{98}. The other coefficients from this specification generally follow sensible patterns: wealth increases, for example, with income, education, and age (Table 2).22

This $6,690 increase in the relative wealth of eligible households is quite large and exceeds the corresponding $4,940 increase—from $3,460 in 1995 to $8,400 in 2001—in the median 401(k) balance of 401(k)-eligible households.23 The increase is also larger than those found in earlier studies (see, for example, Poterba, Venti, and Wise (1995) and Engen and Gale (2000)). Many of those studies, however, were based on 1987-1991 data from the Survey of Income and Program Participation. We might expect the SCF to show a bigger increase than the SIPP because the stock market experienced larger gains over the 1995-2001 period than the 1987-1991 period; because the SCF data provide a more comprehensive measure of wealth than the SIPP; and because the estimates in this study are expressed in 2001 dollars rather than dollars from earlier years.

---

22 To aid convergence, the dependent variable was divided by 10,000 before this specification was run, and the coefficients shown in Table 2 were subsequently multiplied by 10,000.

23 In comparing the $6,690 increase in net financial assets to the $4,490 increase in 401(k) balances, keep in mind that these estimates should not line up exactly because the median is not a linear operator. More precisely, median(\text{net financial assets}) ≠ median(401(k) balances) + median(\text{net financial assets} – 401(k) balances).

*Contributions to Economic Analysis & Policy*
### TABLE 1: Estimated Effect of 401(k) Eligibility on Wealth Accumulation

<table>
<thead>
<tr>
<th>Specification</th>
<th>Eligible</th>
<th>Eligible• Year&lt;sub&gt;98&lt;/sub&gt;</th>
<th>Eligible• Year&lt;sub&gt;01&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Net financial assets</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Levels</td>
<td>4,523***</td>
<td>3,266*</td>
<td>6,690***</td>
</tr>
<tr>
<td></td>
<td>(1,352)</td>
<td>(1,909)</td>
<td>(2,569)</td>
</tr>
<tr>
<td>Inverse hyperbolic sine evaluated at</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median in 1995, $7,797</td>
<td>6,295***</td>
<td>602</td>
<td>1,130</td>
</tr>
<tr>
<td></td>
<td>(790)</td>
<td>(1,002)</td>
<td>(943)</td>
</tr>
<tr>
<td>Median of marginal effects for households in 1995</td>
<td>9,201***</td>
<td>880</td>
<td>1,652</td>
</tr>
<tr>
<td>Mean of marginal effects for households in 1995</td>
<td>51,191***</td>
<td>4,894</td>
<td>9,193</td>
</tr>
<tr>
<td>75&lt;sup&gt;th&lt;/sup&gt; percentile in 1995, $50,000</td>
<td>37,170***</td>
<td>3,557</td>
<td>6,680</td>
</tr>
<tr>
<td></td>
<td>(4,665)</td>
<td>(5,920)</td>
<td>(5,575)</td>
</tr>
<tr>
<td><strong>Net financial assets excluding 401(k) balances</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Levels</td>
<td>436</td>
<td>-120</td>
<td>356</td>
</tr>
<tr>
<td></td>
<td>(764)</td>
<td>(1,077)</td>
<td>(1,162)</td>
</tr>
<tr>
<td>Inverse hyperbolic sine evaluated at</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median in 1995, $7,797</td>
<td>1,363</td>
<td>-773</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>(842)</td>
<td>(1,181)</td>
<td>(1,109)</td>
</tr>
<tr>
<td>Median of marginal effects for households in 1995</td>
<td>1,442</td>
<td>-818</td>
<td>11</td>
</tr>
<tr>
<td>Mean of marginal effects for households in 1995</td>
<td>9,297</td>
<td>-5,272</td>
<td>70</td>
</tr>
<tr>
<td>75&lt;sup&gt;th&lt;/sup&gt; percentile in 1995, $50,000</td>
<td>8,056</td>
<td>-4,568</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>(4,977)</td>
<td>(6,979)</td>
<td>(6,765)</td>
</tr>
<tr>
<td><strong>Net financial assets (households with positive net financial assets)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log</td>
<td>4,167***</td>
<td>579</td>
<td>1,151</td>
</tr>
<tr>
<td></td>
<td>(739)</td>
<td>(1,178)</td>
<td>(1,202)</td>
</tr>
<tr>
<td>Inverse hyperbolic sine evaluated at $7,797</td>
<td>4,146***</td>
<td>992</td>
<td>1,132</td>
</tr>
<tr>
<td></td>
<td>(803)</td>
<td>(1,086)</td>
<td>(1,125)</td>
</tr>
</tbody>
</table>

**NOTES.** Standard errors are bootstrapped with 999 replications and are adjusted for imputation uncertainty. The median and mean of the IHS marginal effects are assumed to have the same statistical significance as the equivalent IHS marginal effects evaluated at $7,797.

**SOURCE.** Survey of Consumer Finances. Data are weighted and expressed in 2001 dollars.

* statistically significant at the 10 percent level
*** statistically significant at the 1 percent level

http://www.bepress.com/bejeap/contributions/vol5/iss1/art20
**TABLE 2: Full Sets of Coefficients**

<table>
<thead>
<tr>
<th></th>
<th>Levels</th>
<th>Inverse Hyperbolic Sine</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>Std. Error</td>
</tr>
<tr>
<td></td>
<td>Coefficient</td>
<td>Std. Error</td>
</tr>
<tr>
<td>Eligible</td>
<td>4,523</td>
<td>1,352</td>
</tr>
<tr>
<td>Eligible • Year=1998</td>
<td>3,266</td>
<td>1,909</td>
</tr>
<tr>
<td>Eligible • Year=2001</td>
<td>6,690</td>
<td>2,569</td>
</tr>
<tr>
<td>Year=1998</td>
<td>2,291</td>
<td>787</td>
</tr>
<tr>
<td>Year=2001</td>
<td>3,616</td>
<td>786</td>
</tr>
<tr>
<td>Income:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$10K-30K</td>
<td>143</td>
<td>789</td>
</tr>
<tr>
<td>$30K-50K</td>
<td>3,098</td>
<td>842</td>
</tr>
<tr>
<td>$50K-80K</td>
<td>16,878</td>
<td>1,455</td>
</tr>
<tr>
<td>$80K-$150K</td>
<td>67,986</td>
<td>4,849</td>
</tr>
<tr>
<td>greater than $150K</td>
<td>262,819</td>
<td>27,606</td>
</tr>
<tr>
<td>Age in 1995:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35-44</td>
<td>5,727</td>
<td>686</td>
</tr>
<tr>
<td>45-54</td>
<td>13,131</td>
<td>1,434</td>
</tr>
<tr>
<td>55-64</td>
<td>25,337</td>
<td>5,178</td>
</tr>
<tr>
<td>Education:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school graduate</td>
<td>1,758</td>
<td>734</td>
</tr>
<tr>
<td>Some college</td>
<td>2,742</td>
<td>945</td>
</tr>
<tr>
<td>College graduate</td>
<td>3,500</td>
<td>1,676</td>
</tr>
<tr>
<td>Graduate school</td>
<td>25,551</td>
<td>5,522</td>
</tr>
<tr>
<td>Marital status:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Divorced</td>
<td>-3,945</td>
<td>691</td>
</tr>
<tr>
<td>Married</td>
<td>-1,784</td>
<td>892</td>
</tr>
<tr>
<td>Two earners</td>
<td>-3,635</td>
<td>1,144</td>
</tr>
<tr>
<td>Not white</td>
<td>-2,554</td>
<td>594</td>
</tr>
<tr>
<td>DB participant</td>
<td>1,115</td>
<td>1,429</td>
</tr>
<tr>
<td>Industry:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mining &amp; Construction</td>
<td>1,886</td>
<td>6,253</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>3,995</td>
<td>6,390</td>
</tr>
<tr>
<td>Trade</td>
<td>979</td>
<td>6,268</td>
</tr>
<tr>
<td>Finance, Insurance, &amp; Real Estate</td>
<td>1,741</td>
<td>6,283</td>
</tr>
<tr>
<td>Services &amp; Public Utilities</td>
<td>-188</td>
<td>6,312</td>
</tr>
<tr>
<td>Government &amp; Military</td>
<td>79</td>
<td>6,356</td>
</tr>
</tbody>
</table>

*Contributions to Economic Analysis & Policy*
### TABLE 2: Full Sets of Coefficients, continued

<table>
<thead>
<tr>
<th>Net Financial Assets</th>
<th>Levels</th>
<th></th>
<th>Inverse Hyperbolic Sine</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>Std. Error</td>
<td>Coefficient</td>
<td>Std. Error</td>
</tr>
<tr>
<td><strong>Occupation:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical and Sales</td>
<td>-4,575</td>
<td>1,308</td>
<td>-1,488</td>
<td>619</td>
</tr>
<tr>
<td>Services</td>
<td>-1,509</td>
<td>1,333</td>
<td>-921</td>
<td>746</td>
</tr>
<tr>
<td>Skilled blue collar</td>
<td>-2,963</td>
<td>1,397</td>
<td>-1,771</td>
<td>665</td>
</tr>
<tr>
<td>Unskilled blue collar</td>
<td>-5,456</td>
<td>1,451</td>
<td>-3,727</td>
<td>747</td>
</tr>
<tr>
<td>Farmers and forestry</td>
<td>2,478</td>
<td>6,474</td>
<td>-65</td>
<td>2,583</td>
</tr>
<tr>
<td><strong>Firm Size:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 – 19</td>
<td>2,156</td>
<td>1,112</td>
<td>1,583</td>
<td>796</td>
</tr>
<tr>
<td>20 – 99</td>
<td>-1,442</td>
<td>990</td>
<td>-594</td>
<td>742</td>
</tr>
<tr>
<td>100 – 499</td>
<td>-573</td>
<td>1,039</td>
<td>-10</td>
<td>869</td>
</tr>
<tr>
<td>500 +</td>
<td>1,543</td>
<td>875</td>
<td>894</td>
<td>663</td>
</tr>
<tr>
<td>Union Coverage</td>
<td>-826</td>
<td>1,001</td>
<td>410</td>
<td>602</td>
</tr>
<tr>
<td>Constant</td>
<td>-2,871</td>
<td>6,494</td>
<td>-118</td>
<td>2,501</td>
</tr>
</tbody>
</table>

**NOTES.** Omitted categories: income less than $10,000; ages 25-34 in 1995; no high school degree; single/widowed; industry = agriculture; occupation = executive and managerial; firm size less than 10 workers. Standard errors bootstrapped with 999 replications and adjusted for imputation uncertainty. IHS coefficients are marginal effects evaluated at median net financial assets in 1995, $7,797.

**SOURCE.** Survey of Consumer Finances. Data are weighted and expressed in 2001 dollars.

In the inverse hyperbolic sine specification, 401(k) eligibility is associated with smaller changes in wealth over time. For ease of comparison with the levels specification, the coefficients are shown as marginal effects evaluated at $7,797, which was median net financial assets in 1995. The coefficients suggest that eligible households held $6,295 more net financial assets than ineligible households in 1995, a difference about comparable to the levels specification (Table 1). However, the relative increases in the wealth of eligible households are much smaller—$602 for the 1995–98 period and $1,130 for the 1995–2001 period—and neither increase is significantly different from zero. The coefficients on household characteristics correlated with high wealth—income above $80,000; graduate education; and age 55 or older—are also damped relative to the levels specification (Table 2).

---

24 These standard errors do not incorporate the uncertainty associated with estimating $\theta$. The statistics literature has not reached a consensus as to whether incorporating this uncertainty is important (Burbidge, Magee, and Robb, 1988, p. 124); the treatment here follows Burbidge, Magee, and Robb (1988).

http://www.bepress.com/bejeap/contributions/vol5/iss1/art20
To establish that the IHS results are robust to the method used to estimate the marginal effects, I also calculate the marginal effect for each household in the 1995 data and take the median and mean of these estimates.\textsuperscript{25} The median marginal effects are comparable in size to the marginal effects evaluated at $7,797: they indicate that eligible households had $9,201 more net financial assets than ineligible households in 1995, and that these assets increased, relative to ineligible households, by $880 over the 1995-1998 period and $1,652 over the 1995-2001 period.

However, the mean marginal effects are larger than the levels estimates: they indicate that eligible households had $51,191 more net financial assets than ineligible households in 1995, and that their assets increased, relative to ineligible households, by $4,894 over the 1995-98 period and $9,193 over the 1995-2001 period. These large estimates reflect the fact that the distribution of net financial assets is skewed and thus the mean can be heavily influenced by outliers. In fact, in order for the levels and IHS coefficients to have the same magnitude, the IHS marginal effect would have to be evaluated at $50,000, about the 75\textsuperscript{th} percentile of the net financial assets distribution, at which value the marginal effect is $6,680 (although still statistically insignificant).

The IHS marginal effects also depend on the chosen value of $\theta$. The estimates in this paper are based on a $\theta$ of 0.0003, the optimal value as estimated by maximum likelihood. However, as $\theta$ approaches 0, the IHS marginal effect approaches the levels regression coefficient (Table 3), whereas for larger values

\begin{table}[h]
\centering
\caption{Eligible\textsuperscript{-}Year\textsubscript{01} Coefficient as a Function of Theta}
\begin{tabular}{lll}
\hline
Theta & Coefficient & Standard Error \\
0.001 & -39 & 1,223 \\
0.0009 & -50 & 1,202 \\
0.0007 & 98 & 1,111 \\
0.0005 & 514 & 1,044 \\
0.0003 & 1,130 & 943 \\
0.0001 & 2,244 & 1,081 \\
0.00005 & 3,451 & 1,369 \\
0.00001 & 6,162 & 2,228 \\
0.000001 & 6,709 & 2,732 \\
\hline
\end{tabular}
\end{table}

\textbf{NOTES.} Standard errors are bootstrapped with 999 replications and are adjusted for imputation uncertainty. Coefficients are marginal effects evaluated at the 1995 median of net financial assets, $7,797.

\textbf{SOURCE.} Survey of Consumer Finances. Data are weighted and expressed in 2001 dollars.

\textsuperscript{25} The median of the marginal effects is not equal to the marginal effect evaluated at the median because the marginal effect function is not monotone.
of $\theta$ the IHS marginal effect goes to zero. Although the marginal effect varies substantially with $\theta$, the range of possible values of $\theta$ for this specification, as estimated by maximum likelihood, is quite small. A bootstrapped 95 percent confidence interval for $\theta$ reveals that $\theta$ likely lies in the interval [0.00025, 0.00032], which corresponds to a possible range of marginal effects from $1,000 to $1,267.

For $\theta=0.0003$, in fact, the IHS specification closely resembles a log specification for most households in the sample. To show this, I limit the sample to households with positive net financial assets. Estimating the log-wealth specification over this sample indicates that eligible households held $4,167 more net financial assets than eligible households in 1995 and that the net financial assets of eligible households, relative to ineligible households, increased by $579 over the 1995–98 period and by $1,151 over the 1995–2001 period (Table 1). Estimating the inverse hyperbolic sine specification over the same sample suggests that eligible households held $4,146 more net financial assets than ineligible households in 1995 and that their assets increased by $992 more over the 1995–98 period and by $1,132 over the 1995–01 period.

**Net financial assets excluding 401(k) balances.** For this wealth measure, I expect the levels and IHS estimates to match more closely, in part because the holdings of eligible and ineligible households are more comparable than for net financial assets. In 1995, median net financial assets excluding 401(k) balances were $9,804 for eligible households and $1,153 for ineligible households. In addition, capital gains on this wealth measure were modest because many 401(k) participants held few stocks outside of their 401(k) accounts. The median 401(k)-eligible stockowner, for example, held 89 percent of his stock wealth within a 401(k) account in 1995.

As anticipated, the levels and IHS specifications yield comparable results. In 1995, eligible households held either $436 (levels specification) or $1,363 (IHS specification, marginal effects) more net financial assets, excluding 401(k) balances, than ineligible households. Over the 1995–98 period, the holdings of eligible households fell by $120 (levels) or $773 (IHS) relative to ineligible households; over the 1995–01 period, their holdings increased by $356 (levels) or $10 (IHS). None of these coefficients are statistically significant.

**Implications.** In the net financial assets specification, the difference between the levels and the IHS estimates suggests that some of the wealth increase of eligible households over this period stemmed from investment returns on their higher initial stock of financial assets. Of course, this analysis does not control for all of the differences between these two groups, such as differences in their taste for saving, that might affect wealth accumulation. Eligible households

---

26 These marginal effects are evaluated at $7,797.
may have started the period with a higher stock of initial wealth because of their greater commitment to saving.

More generally, the results do not answer the question of whether 401(k) eligibility increases saving. Recall that the regression framework introduced in equation (1) offered three tests of whether 401(k) eligibility increased saving: in the net financial assets specification, \( \lambda_{98} \) and \( \lambda_{01} \) (the coefficients on \( \text{eligible} \cdot \text{year}_{98} \) and \( \text{eligible} \cdot \text{year}_{01} \)) should be greater than zero and \( \lambda_{01} > \lambda_{98} \); in the net financial assets excluding 401(k) balances specification, \( \lambda_{98} \) and \( \lambda_{01} \) should be positive, near zero, or negative but small. The results from the net financial assets specification suggest that 401(k) accounts do not increase saving. Although \( \lambda_{01} > \lambda_{98} > 0 \) for both the levels and IHS specification, the IHS estimates are small and not statistically significantly different from zero. However, the results from the net financial assets excluding 401(k) balances specification suggest that 401(k) accounts do increase saving, as \( \lambda_{98} \) and \( \lambda_{01} \) are near zero in both the levels and the IHS specification.

How is it possible that 401(k) accounts both do not increase net financial assets and do not displace other financial assets? Perhaps these accounts displace other assets that are not included in the wealth measure, such as home equity or defined benefit pensions.27 In addition, the tests may not have enough statistical power to determine the effect of 401(k) eligibility on saving. Wealth data, including the SCF, are notoriously noisy, and 401(k) contributions were a relatively small part of the change in wealth during this time period. Only eighteen percent of the change in aggregate defined contribution pension balances from 1995 to 2001, as measured in the Flow of Funds Accounts, came from contributions.28

**CONCLUSION**

This paper describes how the inverse hyperbolic sine transformation can help resolve a problem commonly encountered when estimating the effects of tax incentives on saving. Specifically, researchers may want to evaluate the effects of these incentives by comparing changes in the levels of wealth of households eligible for and ineligible for the incentives. If eligible households begin the period with more wealth than ineligible households, and if asset values change significantly over the period, the outsized asset gains or losses that eligible

---

27 Unfortunately I cannot determine definitively with the SCF data whether 401(k) accounts displace nonfinancial assets, in part because these assets are difficult to measure. Some specifications suggest that 401(k) balances displace home equity, but this result is not consistent across all specifications and often is statistically insignificant.

28 Paul Smith of the Federal Reserve Board’s Flow of Funds section graciously provided this estimate.
households realize on their higher stock of assets can dominate any effects of the
tax incentives on saving behavior.

The inverse hyperbolic sine addresses this problem by allowing
researchers to estimate a percentage change specification on data that include
nonpositive values. I show that when estimating the relative change in net
financial assets—a wealth measure for which both differences in initial wealth
and changes in asset values were significant—the levels and IHS specifications
give quite different answers. In the case of the net financial assets excluding
401(k) balances—where these conditions do not hold—the two specifications
yield similar answers.

More generally, the inverse hyperbolic sine can help any researcher who
wants to estimate a percentage change specification without dropping nonpositive
data values. This transformation also lets researchers ascertain whether a levels
or percentage change specification provides the best fit to their data. By damping
outliers, the IHS, like the logarithm, can also ameliorate heteroskedasticity.
Researchers who study data that take on economically significant negative values,
such as wealth, corporate profits, self-employment income, or net investment,
may find this transformation useful.

REFERENCES

Aizcorbe, Ana M., Arthur B. Kennickell, and Kevin B. Moore, “Recent Changes
in U.S. Family Finances: Evidence from the 1998 and 2001 Survey of

Attanasio, Orazio, and Thomas DeLeire, “The Effect of Individual Retirement
Accounts on Household Consumption and National Saving,” Economic

Benjamin, Daniel J., “Does 401(k) Eligibility Increase Saving? Evidence from
Propensity Score Subclassification,” Journal of Public Economics, 87(5-
6), May 2003, 1259-90.

Bernheim, B. Douglas, “Taxation and Saving,” Handbook of Public Economics
Volume 3, Alan J. Auerbach and Martin Feldstein, eds., Amsterdam:

Burbidge, John B., Lonnie Magee, and A. Leslie Robb, “Alternative
Transformations to Handle Extreme Values of the Dependent Variable,”

http://www.bepress.com/bejeap/contributions/vol5/iss1/art20


Ma, Jennifer, “Education Saving Incentives and Household Saving: Evidence from the 2000 TIAA-CREF Survey of Participant Finances,” *College


http://www.bepress.com/bejeap/contributions/vol5/iss1/art20