Consumption Smoothing among Working-Class American Families before Social Insurance

by

John A. James

Michael G. Palumbo

Mark Thomas

Abstract: This paper examines whether the saving decisions of a large sample of working-class American families around the turn of the twentieth century are consistent with consumption smoothing tendencies in the spirit of the permanent income hypothesis. We develop two econometric models to decompose reported annual incomes from micro-data into expected and unexpected components, then we estimate marginal propensities to save out of each component of income. The two methodologies deliver similar regression estimates and reveal empirical patterns consistent to those reported in other recent research based on quite different contemporary household data. Marginal propensities to save out of unexpected income shocks are large relative to propensities based on expected income movements, though the former lie much below one and the latter much above zero. While these data reject strict parameterizations of the permanent income hypothesis, we nonetheless conclude that families' saving decisions in the historical period look quite "modern."

JEL Classification Codes: D91, N31, E21.
Keywords: Unemployment risk, permanent income hypothesis, precautionary saving.

Affiliations and Acknowledgments: James is in the Department of Economics at the University of Virginia, Charlottesville, VA 22903 (jaj8y@virginia.edu); Palumbo is at the Federal Reserve Board, Mail Stop 80, Washington, D.C. 20551 (mpalumbo@frb.gov); and Thomas is in the Department of History at the University of Virginia (mark.thomas@virginia.edu). We appreciate discussing this research with Martin Browning, Thomas Crossley, Chinhui Juhn, David Papell, Jonathan Parker and Luigi Pistaferri. Also, seminar audiences at the American Economic Association, the Federal Reserve Board, McMaster University and York University provided valuable comments. The views presented are solely those of the authors and are not necessarily shared by the Federal Reserve Board or its staff.

May 1999

Consumption Smoothing among Working-Class American Families before Social Insurance

Around the turn of the twentieth century, many Americans lived and worked in an environment of considerable economic uncertainty. Industrial accidents presented significant risks to many (see Kantor and Fishback, 1996), as did illnesses, but even more pervasive was the risk of unemployment. Unemployment was much more widespread during the period before World War I than it has been since World War II. Not only was the natural rate of unemployment higher, but so too was the cyclical sensitivity of unemployment (James and Thomas, 1996). Moreover, the incidence of unemployment was more widespread, implying that a greater proportion of workers had need for precautionary action than today. Unemployment in this historical period had its predictable elements -- the availability of work followed strong seasonal influences, for one thing -- but loss of work also resulted from much less predictable factors. Business cycle downturns during the late nineteenth and early twentieth centuries were on average more serious than they had been before or have been since. The severity of business cycles had increased dramatically from the period before the Civil War to the one after (James, 1993) -- indicating an increasing need for precautionary behavior -- and this in a period before the rise of governmental institutions designed to take the sting out of unemployment spells. Workers in the late nineteenth century were essentially dependent on their own devices to combat income uncertainty. It is perhaps ironic that the expansion of social insurance after World War II coincided with a moderation of unemployment volatility.[1]

Alexander Keyssar, in his well-known study of unemployment in Massachusetts, observes that employment for workers in this period was "chronically unsteady" (1986, p. 59). Even within the business cycle he stresses the great diversity of individual experience: "The incidence of joblessness during depressions was always checkered, erratic, variegated." Moreover, "a majority of the working class found that the threat of unemployment remained palpable even when business was good." (1986, pp. 55, 58). Substantial negative shocks to household income therefore would have been common and, especially for lower income families, presented potential economic disasters. Workers faced such risks without assistance from any sort of public safety net -- no unemployment compensation, no worker's compensation for accidents, no sickness coverage. Similarly, few workers could access formal credit markets during this time to enable smoothing adverse income shocks by taking temporary loans.

[1] Long run changes in the degree of volatility of national output and unemployment in the U.S., of course, has been a controversial issue (see Romer, 1986; Weir, 1992). However, James and Thomas (1996) document a decrease in the cyclical response of unemployment from the pre-World War I period to that post-World War II.

Commercial banks, influenced at least in theory by the real bills doctrine, limited themselves to commercial rather than personal loans, while mutual and stock savings banks were less than ubiquitous.

A similar environment of unpredictable incomes and weak formal credit markets exists for low-income households in developing countries today. Nonetheless, a number of recent studies have found that, even without access to formal insurance and credit markets, such households have generally been able to smooth their expenditures in the face of large and frequent income shocks (e.g., Deaton 1990; Paxson 1992; Townsend 1995). However, most previous research on how households in developing countries cope with economic risks study families who rely heavily on agricultural production for their earnings and thus face different circumstances than did American industrial workers a hundred years ago. Crop diversification and the maintenance of buffer stocks of commodities were not options open to the non-farm working class in America, on whom we focus in this paper.

Does it follow, therefore, that American working-class families responded to uncertain earnings prospects through different private saving patterns than has been observed among peasant farm households today? Or does it mean that workers found means other than private saving to smooth consumption in the face of volatile and unpredictable incomes? We address these themes using a two-pronged approach. Most of the paper reports on an econometric analysis of the role played by private saving as a potential means for buffering volatile income experiences by working-class families a hundred years ago. A second shorter historical section then analyzes the institutional environment surrounding families of the time to complete the picture. Our historical survey reports on the availability of consumer credit, insurance coverage and pawnbroking services, all of which might be expected to have enhanced the ability of working-class Americans during this period to protect their expenditures from the short-term income disturbances to which they were vulnerable.

The plan of the paper goes as follows. The first section describes the most comprehensive micro-level, primary-source database assembled to date on saving, income and employment covering American workers surveyed during twenty-plus years around the turn of the century. By merging information from thirty cross-section surveys based on nearly identically-worded questionnaires, our database includes information from more than 32,000 working-class families interviewed between 1884 and 1909. In the second section, we describe an econometric methodology, similar to that proposed by Paxson (1992), for decomposing annual income realizations into predictable and unexpected components for each worker in the sample.

The third and fourth sections contain the primary empirical results of the paper based on two sets of estimated marginal propensities to save out of predictable and unexpected income components. We first pursue an econometric methodology which exploits variation in unemployment during the previous year at the worker level to identify the marginal propensity to save from unexpected income realizations. This method carries some intuitive appeal, but is subject to fairly standard endogeneity criticisms. Our second econometric approach involves aggregating the micro data up to "groups" of workers, based on shared characteristics, then identifying the marginal propensity to save out of unexpected income realizations from "business cycle" variation in group-average outcomes across time periods. The two disparate empirical methodologies yield generally consistent results and, in particular, econometric estimates from the aggregate-level analysis show no evidence that potential endogeneity is responsible for the micro-level regression results.[2]

Both empirical analyses present strong empirical evidence that working-class American households used their own saving to smooth consumption in the face of volatile employment circumstances during the late nineteenth and early twentieth centuries, much as contemporary American and European families, as well as farmers in developing countries appear to do. As almost all recent studies find as well, the regression coefficients for household saving reject strict specifications of the permanent income hypothesis, but seem in line with what one might expect among precautionary or buffer-stock saving behavior. We conclude, therefore, that the broad saving patterns in this unique historical micro-level database fit "modern" intertemporal behavior, in this sense.[3]

As mentioned, the fifth section of the paper presents some historical evidence on the availability of financial institutions that would have accomodated consumption smoothing by American workers during this period through means other than their own private saving. We argue that these alternative mechanisms which might have used to complement private saving as a means to buffer household expenditures were in fact quite limited. Finally the paper concludes by briefly summarizing our empirical findings and comparing them with results and conclusions from previous related research.

I. Saving and Income Data from a Series of Worker Surveys

[2] We discuss these issues in detail below. Our aggregate-level, or time-series, approach, but not our micro-level regression, is appropriate if workers differ in terms of their unobserved tendencies to be employed and to save resources for future uncertain contingencies. Despite the intuitive appeal of such a hypothesis, we uncover no support for the endogeneity it implies in these historical data.

[3] Note, we leave for future research the more ambitious, and necessarily complex, question of whether these micro-level saving data are completely consistent with intertemporal optimization under rational expectations, given the magnitude of income risks apparently facing this sample of American workers.

State-level bureaus of labor statistics published more than one hundred surveys of wage-earning workers during the late nineteenth and early twentieth centuries. The surveys focused on economic conditions facing American workers, employed mostly in the non-farm sector, and the living conditions of their families. Some surveys concentrated on covering workers employed in specific industries (vehicle manufacture; iron production; furniture), actually holding interviews at the workplace (the Michigan model); others polled representative samples of workers across industries by mail (the Kansas model). Although the information collected varied from survey to survey, most followed a common model using similarly worded or identical questions, following the example developed by Carroll D. Wright in Massachusetts, who began surveying workers during the early 1860's. Individual survey responses were invariably published in full, without editorial embellishment or alteration, once accuracy and consistency were checked. Each survey covered a different cross-section of workers; none followed the experiences of individual workers or families across more than one year. It is not possible therefore to construct a true panel from these sources; our database consists of merged independent cross-section surveys.[4] As Table A-1 indicates, the database includes 32,150 workers from 8 states interviewed during 22 different years from 1884 through 1909. Note, however, that nearly all (97 percent) of the families in our sample reside in Kansas or Michigan; most surveys used here (71 percent) were conducted during the 1890's.[5]

We limit our analysis in this paper to those thirty surveys which report information on income, saving and/or expenditures and days of unemployment during the previous year. Additionally, we know the skill level for each worker, as well as his age, industry, state of residence and the year of the survey.[6] Income in this paper refers to annual family income, combining income from all sources. A few surveys report labor earnings disaggregated by earner (or, at least, separately for primary wage-earners and all others in the family), but most do not.[7] The database employs two different measures

---

[4] We discuss the dataset in more detail elsewhere (James, Palumbo, and Thomas 1996), so here we shall be brief.

[5] In fact, we have repeated all of the paper's empirical analysis using only data from Michigan and Kansas. None of our results are affected significantly by focusing exclusively on this subsample.

[6] Our working database excludes the few women surveyed, as well as the few men younger than 15 years of age or older than 75.

[7] None of the surveys separate interest income from labor income, but the former category is unlikely to be a major contributor among workers included in this sample.

of annual saving.[8] Some households reported last year's saving directly (answering a question like "How much of your income did you save last year?"); others recorded total family expenditures, from which saving may be calculated as the residual from annual income.[9] We designate the first variable as "reported saving," and the second as "calculated saving." We apply all the empirical analyses in the paper to both measures of household saving.

Figures 1 and 2 show empirical density and distribution functions for reported and calculated saving based on the pooled cross-section survey data. The figures show clear differences between the two distributions. Surveys that measured saving directly only recorded additional money set aside (positive values for reported saving); otherwise, entries were left blank. After analysis of the data, we concluded that 'no response' generally indicated zero (or negative) saving.[10] Thus, Figure 1 shows some possible effects of left-censoring at zero – a large spike in the distribution function at the censoring point— as well as other smaller spikes at "round" numbers.[11] Calculated saving, on the other hand, takes both positive and negative values. Thus, Figure 2 shows a less skewed distribution of saving, and a smaller spike occurring exactly at zero dollars per year. Appendix A reports a detailed investigation into the differences between the two saving series. Among a subsample of families for whom we observe both reported and calculated saving, the two variables are strongly linearly related. Thus, the divergence between the distributions shown in Figures 1 and 2 turns out to

---

[8] Note that the surveys clearly asked respondents about the annual flow of saving out of income during the previous year, rather than about their accumulated stocks of assets or net worth. Some of the respondents, naturally, owned their homes. During this period, home mortgages typically involved short-term contracts in which only interest payments occurred during the loan's duration with a balloon payment of principal due at maturity. Thus, paying down the loan principal at maturity must have accomplished through prior saving and, we have argued (James, Palumbo and Thomas 1996), our saving measures seem likely to include changes in home equity.

[9] Direct savings were recorded by 26,112 families, while savings were calculated as a residual for a further 11,946 families (see Table A-1).

[10] One survey, covering vehicle workers in Michigan during 1896, recorded three responses to the question "amount saved last year?" – a positive value, "none," and a blank response. Of the 2,787 survey respondents who do not report positive saving (out of 3,776 total observations), 2,576 explicitly answer "none"; only 211 have a blank response.

[11] For consistency, we artificially censored the few observations for which we observe negative values for reported saving. These come from a single Kansas survey in which respondents answered "Did you save or run a deficit last year?"

have virtually no consequence for our empirical analyses. As we report below, Tobit equations estimated using reported saving yield nearly identical slope coefficients to OLS regressions based on calculated saving, though the estimated intercepts differ between the specifications.

Table 1 presents mean income, mean consumption and the variance of expenditure relative to the variance of income for workers in our dataset who report information on annual saving directly (grouped by age, by skill and by industry). Similar information is shown in Table 2 for those workers who directly report income and expenditures (i.e., for those whose savings have been calculated by us). Both tables clearly indicate the tendency for variability in family income to exceed variability in expenditure, regardless of classification. In the subsample for which calculated saving is available (Table 2) the cell variance of consumption is generally at least 30 percent less than that of income; where saving reported directly (Table 1), the difference is somewhat smaller -- expenditures vary about 20 percent less than incomes, on average.

An interesting pattern follows from comparing mean income and consumption between savers and nonsavers within each cell. In virtually every cell of Tables 1 and 2, savers earn higher incomes than nonsavers, but average consumption levels between the two groups are extremely similar. The few cells in which consumption differences are relatively large all suffer from relatively small cell sizes. Among families in the calculated saving subsample, mean consumption levels differ only by 8 percent between savers and nonsavers, but income levels are about 34 percent greater among savers on average. This striking result suggests that saving might have responded largely to income "surprises" among these families, thereby motivating the empirical strategies to be described next.

II. Estimating Predictable and Unexpected Components of Income from Annual Reported Unemployment

A key implication from theories of household saving based on intertemporal optimization is that marginal propensities to save (mps; or to consume, mpc) differ by composition of income. The empirical strategy for examining saving behavior employed in this paper involves making explicit comparisons of the marginal propensities to save out of predicted and unexpected (transitory) income to discern motives for saving. “Keynesian-saving” families, whose spending simply is a function of current income, ought to have marginal propensities to save that apply equally to all components of annual income. On the other hand, if families are guided by a “certainty equivalence” decision rule (or, according to the “permanent income hypothesis” in Deaton’s (1992) terminology), then predictable differences in income levels will not affect observed saving levels, but unexpected income shocks will affect saving decisions dollar-for-dollar.[12] Finally, recent theoretical models based on intertemporal optimization with unpredictable family incomes and "prudence" (a characteristic of household utility; see Carroll, 1997; or Deaton, 1992) or liquidity constraints (Deaton, 1991) imply small, but nonzero, marginal propensities to save out of predictable incomes, and marginal propensities close to one out of unexpected income.[13]

This approach to econometric analysis of household saving behavior thus requires realized family income each year to be decomposed into its predictable and unexpected parts. Following Paxson (1992), rather than estimating transitory income simply as a residual from a regression equation for annual income on predictable family and worker characteristics, we use survey information to measure it more directly. In her study of saving behavior among Thai rice farmers, Paxson uses deviations in rainfall from historical averages to measure unexpected income shocks for small geographic regions in Thailand. Our application focuses on shocks to time spent out of work among primary earners in working-class families. Unexpected days lost from work would not have directly affected family expenditures, but would have translated into important income shocks to families in our database, which should flow into changes in savings levels, according to modern theory. A deviation in reported workdays lost from its predicted value, therefore, provides a measure of unexpected income shock realized by each family in our database, in a fashion analogous to that produced by variation in annual rainfall among Paxson's sample of Thai rice farmers.

Our measure of time out of work, which we call "workdays lost", actually measures nonemployment (deviations from full-time employment) over the course of the entire survey year, as reported by each respondent and possibly occurring for a variety of reasons (low-frequency job loss, high-frequency inabilities to find work, accidents, sicknesses, etc.). Column 1 of Table 3 summarizes the distribution of annual workdays lost among different categories of respondents. The average worker missed 37 days of work during the previous year; the median length of time lost being 18 days. Twenty-five percent of respondents report not missing any work during the previous year; another twenty-five percent report missing more than fifty workdays last year. Clearly, substantial variation

---

[12] A concise, formal derivation for optimal consumption and saving rules under certainty equivalence, or the permanent income hypothesis, can be found in Pistaferri (1998); Paxson (1992) includes an informal discussion of similar results.

[13] This statement is made somewhat speculatively. We are not aware of regressions based on simulated buffer stock or precautionary saving models (other than some of our own) that specifically highlight theory-based coefficients directly comparable to those we estimate in this paper. In other related work, we have begun to examine these using contemporary model specifications calibrated to describe our historical environment (James, Palumbo and Thomas, 1999).

in workdays lost exists in the microdata. In this paper, we essentially ask how variation in lost workdays contributes to variation in family or group saving decisions.

Decomposing annual income into its predictable and unexpected parts requires us first to decompose annual workdays lost for each survey respondent.[14] We first estimate a regression for annual workdays lost as a function of each respondent's age, state of residence interacted with survey year, and skill category interacted with industry of employment.[15] Using the regression-fitted value to estimate the predictable number of workdays lost for each observation, the regression residual estimates the unexpected shock to employment experienced by each worker during the previous year. Our specification is quite flexible and, because all the explanatory variables are categorical indicators, the regression effectively defines the predictable number of workdays lost to be the average among all workers of a particular type. We assign workers to types, or cells, defined by four age categories, twenty-one survey groups (state-by-survey year combinations) and twenty-two occupational skill-by-industry categories.[16] The adjusted R^2 of the regression equation for reported workdays lost during the survey year on the categorical indicator variables for worker type, which is estimated using all 32,150 observations in the sample, is 0.12. Rather than report all the estimated regression coefficients, which are cumbersome to interpret, column (1) of Table 3 shows average workdays lost during the previous year by some of the categories of worker types. The regression results show some clear differences in average workdays lost among different groups of workers.

The unexpected component of annual workdays lost is defined as each individual survey respondent's deviation in reported workdays lost during the previous year from the average workdays lost among members of his type. The second and third columns of Table 3 show estimated dispersion in unexpected workdays lost during the previous year obtained from this procedure, as measured by

---

[14] The appendix to the paper includes an algebraic representation of the model described verbally here.

[15] We experimented with numerous alternative regression specifications, some of which include a national, time-varying business cycle index (and its interactions with other variables) instead of year dummies. This paper's results are based on a parsimonious specification described in the text. The alternative models produce very similar results.

[16] For comparison, Paxson's "cells" are defined for individual farmers by their geographic proximity to the nearest of many weather stations located throughout Thailand. Then, it is as if Paxson has a long time-series of rainfall data on which she estimates a regression of annual rainfall on weather station dummy variables. The regression residuals then provide her measure of unexpected annual rainfall, which leads to an estimate of unexpected farm income for each sample observation, following the procedure we describe next.

the interquartile range by (some of the) worker types. Note that, by construction, unexpected workdays lost must average zero for each worker type.[17] Figure 3 shows the density and distribution functions for unexpected workdays lost realized by workers in our sample estimated according to these procedures.

Having in hand an estimate of unexpected workdays lost, we now must translate that variable into a measure of unexpected income during the previous year for each family in the sample. This involves estimating a regression for annual family income on unexpected workdays lost by its primary wage-earner during the previous year and the same set of explanatory variables used in the workdays lost equation to measure predictable income movements. The income regression, again estimated using the entire sample of 32,150 families, yields an adjusted R^2 equal to 0.43.[18] Then, unexpected annual family income is calculated as the product of estimated unexpected workdays lost during the previous year and its estimated coefficient from the family income regression (-1.7214 with a t statistic of -78.40). It seems noteworthy that our procedure produces an estimated "price" for the primary wage-earner missing a day of work extremely close to the average daily wage directly reported in the surveys, $1.92. Missing workdays experienced by primary wage-earners likely represented an important exposure to economic risk among these families -- wages lost by the primary wage-earners do not appear to have been readily made up through additional wages earned by other family members or through other sources of family income.

Predictable family income for each sample household is estimated using the fitted value from the age-, survey- and skill-by-industry indicator variables (and their estimated coefficients) in the family income regression. Column (1) of Table 4 reports average family income, which equals average predicted income by definition, across some of the worker types defined in the regression. Unexpected income for each family is estimated as the product of the regression coefficient -$1.72 and each worker's unexpected workdays lost during the previous year. The distribution of unexpected income

---

[17] For parsimony, we construct Tables 3 and 4 (discussed next) by collapsing the 22 skill-by-industry cells down into four skill categories and five industries. Also, we omit summary statistics by state and year categories simply to conserve space.

[18] Including unexpected workdays lost along with the 44 cell-indicator variables in the annual family income regression produces an R^2 equal to 0.43. Omitting unexpected workdays lost reduces the explained variation in family income to 0.32. Finally, note that we include the same cell indicator variables in the workdays lost regression and in the annual income regression. This means that the estimated coefficient on unexpected workdays lost in the income equation is identical to the coefficient estimate on actual workdays lost (not the residual from a first-stage regression), if we included that variable in the income equation instead.

shocks by some of the worker types is summarized in columns (2) and (3) of Table 4. Finally, the difference between income actually reported by each family in the survey and estimated predicted and unexpected income levels can be computed. Following Paxson (1992), we call this residual "unexplained annual income."

To summarize our empirical methods to this point, respondents in our sample are assumed to have used average incomes earned by other families "like themselves" -- a group defined by their age, state of residence and survey year, as well as their occupational skill and industry classification -- to predict their annual incomes. The estimated coefficient on unexpected workdays lost in the family income equation allows us to derive a dollar-measure of unexpected income. These unexpected income shocks, therefore, average zero across all workers of a given type – or within each cell -- but vary among the individuals within each cell. We use predictable income differences and unexpected income shocks as regressors in an equation for family saving decisions, as described next.[19]

III. Explaining Household Saving with Predicted and Unexpected Income Using Micro-Level Variation in Annual Unemployment

We report results in which household saving is taken to be a linear function of the components of realized annual income:[20]

$$S_{i(t)} = \alpha_0 + \alpha_1 \hat{Y}^P_{i(t)} + \alpha_2 \hat{Y}^T_{i(t)} + \alpha_3 \hat{Y}^U_{i(t)} + \varepsilon^S_{i(t)}, \tag{1}$$

where the subscript, i(t), denotes an observation on family i surveyed during year t and the explanatory variables, respectively, are predicted income, unexpected income and unexplained income components, as estimated according to the methods described in section II.[21] According to a strict certainty

---

[19] As mentioned in the introduction and detailed below, we also develop and apply a second econometric approach using "group-level" estimates of expected and unexpected income realizations and saving regressions.

[20] Experimentation with various nonlinear functions (income categories; splines; quadratic/cubic terms) did not indicate substantial or significant departures from linearity in these data. Paxson (1992) uses time-series data on regional rainfall to estimate the variance of annual income among her sample of Thai farmers, which she includes in her saving equation (1). Our cross-section data preclude any such variable from being estimated and included in the regression analysis.

[21] The subscript notation, i(t), is used in the literature to differentiate data like ours, which contains multiple, independent cross-section surveys, from true panel data -- repeated observations over time from a fixed cross-section of families (usually denoted by the subscripts "it").

equivalence, or permanent income, model of intertemporal household allocation, the marginal propensity to save out of predicted income, α_1, should be zero, while the marginal propensity to save out of unexpected income shocks, α_2, should be one. In this section, we describe regressions for the respondent-level survey data; in the subsequent section, we present results based on panel data for grouped observations for reasons to be described below.

Tables 5 and 6 present the primary estimation results for several model specifications. Table 5 shows results based on Tobit equations using the left-censored (at zero), reported saving as the dependent variable; Table 6 contains OLS estimates when calculated saving is the dependent variable. The first column in each table presents the estimated results from the baseline specification; columns (2) through (4) report results from a few of the many alternative model specifications with which we experimented. Tables 5 and 6 report consistent standard error estimates based on a complete resampling bootstrap procedure using 200 replications of the three-stage estimation procedure.[22] The tables reveal nearly identical marginal propensities to save between the Tobit equations based on reported saving and the OLS estimates based on calculated saving. Furthermore, our basic results are robust with respect to many alternative model specifications and sample selection criteria.

The regressions here show that both predictable and unexpected components of annual income influence household annual saving, but not with the same marginal effects. Marginal propensities to save are much larger (almost twice the magnitude) for unexpected movements in annual income than for predictable income changes. This evidence is inconsistent with the notion of most working-class American families a century ago living "hand-to-mouth" (i.e., $\alpha_1=0$, $\alpha_2=0$), spending their current incomes each year, or even following simple "Keynesian" consumption rules ($\alpha_1=\alpha_2$) . However, at the same time neither are the estimates consistent with a strict certainty equivalence model of intertemporal optimization. Predicted income changes explain a substantial proportion of the variation in annual saving levels according to our survey data, which would not be the case under strict a "permanent income" or certainty equivalence hypothesis ($\alpha_1=0$, $\alpha_2=1$). Furthermore, the marginal propensity to save out of transitory income is smaller than one, its value under a strict certainty

---

[22] We repeatedly estimated all three equations of our empirical model— workdays lost, family income, and saving as a function of expected and unexpected income components — using 200 bootstrap random samples, then constructed standard errors for the saving equation regression coefficients by computing standard deviations across the replicated samples. OLS standard errors are incorrect because of the presence of constructed variables (the three income components) in the saving regression equations.

equivalence behavioral model. Since unexplained income is a mixture of permanent and transitory components, its estimated propensity to save should be a combination of those applied to predicted and unexpected shocks to income. In every case, estimates of α_3 lie between those of α_1 and α_2.

Columns 2 through 4 demonstrate the insensitivity of our basic empirical results to several alternative specifications. The estimates in column 2 come from a specification designed to reduce the influence of the Michigan data relative to those from other states. Given the dominance of Michigan surveys in our pooled dataset, it is important to determine whether our results are driven by unrepresentative behavior of upper mid-western households. Moreover, the implicit weighting scheme imposed by our database is completely arbitrary – it arises not because so many working-class Americans during this period lived in Michigan, but rather because the Michigan state government chose to allocate the most resources toward gathering information about family saving. To counter any possible bias, we reweighted our original sample to reduce Michigan's influence on the parameter estimates. Accordingly, we construct a new set of survey weights to generate hypothetical coverage of 10,000 respondents in each survey and then reestimate the saving equations by Tobit and OLS. As Tables 5 and 6 show, our basic empirical results are not driven solely by the behavior of Michigan respondents. In fact, column 2 shows slightly "better" results, from the permanent income perspective, among non-Michigan respondents than among Michiganers. In the third column of Tables 5 and 6, we effectively employ a third set of sample weights, by estimating the saving equations using only respondents from Kansas and Michigan, which together provide more than 95 percent of our original sample. As might be expected, this revision does not alter the basic parameter estimates relative to the baseline case.

Parameter estimates are remarkably similar, regardless of whether saving is measured by reported or calculated saving as well as whether the smaller state surveys are given more or less weight. The results from all the specifications reported in Tables 5 and 6 indicate that workers saved a large proportion of transitory income, generally about 50 to 60 percent (point estimates range between 0.4771 and 0.5770). All the estimated propensities to save out of transitory income are much larger than zero, and all are sufficiently precise to reject the hypothesis that they equal one. In every case, the saving propensities out of transitory income (α_2) are much larger than those out of permanent income (α_1). The latter are much smaller, ranging between .1738 and .3422, but clearly non-zero by any conventional level of statistical significance. Working-class families at the turn of the century seemed generally to have lived neither "hand-to-mouth" nor, on the other hand, by the dictates of a strict version of the permanent income hypothesis.

To allow for the possibility that longer-term life-cycle factors may also have had an influence on saving other than through permanent and transitory income, we include a quadratic term in age of the household head in the regression equation. The results are shown in the fourth column of Tables 4 and 5. Note, most importantly, that the inclusion of the age and age squared variables has little effect on the estimates of the other variables. Both equations show saving increases with age, other things equal, although the shape of the age functions differ.

We also have investigated some additional specifications of the saving equation, the results of which have been omitted from the tables to conserve space. First, by interacting dummy variables for three age categories with all three income components, we estimated saving equations with age-specific values of α_1, α_2 and α_3. The results indicate quite different reported-saving equations for respondents aged less than 25 years compared to older respondents.[23] Among the youngest group of respondents, marginal propensities to save out of expected income were nearly as large as the marginal propensities to save out of unanticipated income. These results were not replicated in regressions based on calculated saving -- which raises questions about the robustness of age-specific savings behavior. Second, we explored saving equations in which skill indicators or industry indicators were interacted with income components. These regression specifications, however, did not yield substantially different behaviors among different groups of respondents.

IV. Propensities to Save Estimated from Aggregated Data Using Time-Series Variation in Annual Unemployment

The regressions described so far would produce biased estimates of α_1 and α_2 if regression errors for family saving are correlated with expected and unexpected income components at the respondent level. A potential problem arises because expected and unexpected components of income are variables constructed from regression fitted values, since the surveys do not provide direct

[23] Marginal propensities to save were found not to differ among respondents between 25 and 50 years of age and those older than 50 years.

information on these values. This raises the vexed issue of unobserved heterogeneity. What if, for example, respondents knew more about their own work prospects than we attribute to them *and* if those who face the worst work prospects differ in their saving propensities systematically from those with the best work opportunities? In that case, our regressions would have incorrectly attributed the impacts of expected and unexpected income levels to family saving decisions.

To be concrete, consider the empirical implications for respondent-level regressions of the following type of unobserved heterogeneity with respect to labor supply and family saving behavior. Suppose, for simplicity, that two types of workers belong to our historical sample. Type A respondents are hard-working and foresighted – they tend both to work a lot of days during the year and to save relatively large portions of their incomes. On the other hand, Type B respondents are less eager workers and are myopic. Further, suppose all groups of workers (defined as combinations of age, skill, industry, state and year categories in reference to our "workdays lost" and income regressions) contain some respondents of both types. Type A respondents can be expected to report below- (group) average workdays missed during the previous year, while Type B respondents report above-average workdays lost. Additionally, Type A respondents would save more than average, while those of Type B would pull the average down.

Our regression procedure, however, would have estimated identical expected incomes for Type A and B workers who belong to the same age-skill-industry-state-year group, even though the workers themselves in each category expected different incomes, given identical labor demand conditions. We would overestimate expected income among Type B workers and underestimate it among those of Type A. Consequently, our procedures would overestimate unexpected income among Type A workers (the estimates would be too large and positive, on average) and underestimate it among Type B workers (theirs would tend to be "too negative").[24] Incorrect estimation of expected and unexpected incomes due to unobserved heterogeneity lead respondent-level regressions, such as those reported in Tables 5 and 6, to underestimate α_1 and to overestimate α_2, according to this example.

Access to panel data, with multiple observations of lost workdays, income and saving for each respondent, would allow models with individual-specific effects to be estimated as a solution to this potential econometric problem. However, such panel data are unavailable to us, so we instead use a

[24] Recall, we calculate unexpected income as the product of the deviation between reported and predicted workdays lost during the previous year and a "price" coefficient ($-1.72 according to our income equation). Expected income is estimated to be the average annual income for all workers in the respondent's group – defined by age-skill-industry-state-year categories.

different approach to investigate the empirical relevance of potential endogeneity. Our alternative procedure is based on the recognition that, although each worker in our database contributes only a single observation to the sample, we can construct multiple observations (several years, potentially) for *groups* of workers. Then, under the condition that each group contains the same distribution of workers across types, we estimate group-level regressions of saving on expected and unexpected income and compare parameter estimates to those already presented.

Group-level regressions, described in detail next, do not reveal larger estimates of α_1 or smaller estimates of α_2. These alternative specifications, therefore, do not support the concern than unobserved heterogeneity with respect to labor supply and saving behavior (of the particular type described above) generates incorrect inferences from respondent-level regression analysis.

The procedure used to estimate group-level regressions is as follows. For each respondent in our repeated cross-section database for whom we observed calculated saving[25] (N=11,948), we define a cell indicator based on the intersection of the following categorical variables: survey year, occupational skill and industry of employment.[26] This produces 188 cells of data – about 7.8 annual observations for each of the 24 skill-by-industry groups. We then exclude cells containing fewer than 20 respondents and groups with fewer than 4 years of annual observations available.[27] After this, 76 cells remain for analysis – about 11.8 annual observations on 8 skill-by-industry groups – covering 10,678 original sample respondents.

We use the grouped data to estimate a regression similar to (1), estimating cell-level average saving as a linear function of expected and unexpected (and unexplained) income. Expected income is estimated for the group as the average annual income across all available survey years. Unexpected income for each group during each year is estimated to be the product of unexpected workdays lost for the group each year and the "price" of a lost workday, -\$1.72. Unexpected workdays lost is

---

[25] The group-level regressions specify cell-level, average saving as the dependent variable. Our group-level analysis, therefore, focuses on the calculated-saving subsample. Simply averaging the heavily left-censored reported-saving variable across families is inappropriate. Since the micro-level regressions are so similar using the two different saving measures, we are confident of the robustness of these results based only on calculated saving.

[26] As discussed below, we consider several other group definitions, as well as skill-by-industry.

[27] Very small cells are dropped to lessen the impact of outlying observations on the group-level results; groups contributing very few annual observations are dropped because intertemporal variation is what will identify the marginal propensity to save out of unexpected income in the group-level regressions.

estimated as the difference between the average number of workdays lost among members of the group during the current year and the average number of workdays lost across all available survey years.[28]

The group-level regressions identify the extent to which annual deviations in family income arising from variations in workdays lost at a yearly frequency explain variation in annual saving decisions at the group level from their group-specific average across all years. Thus, identification of the marginal propensity to save unexpected income realizations comes from variation at the "business cycle" frequency according to this approach, rather than from variation across families as in Section III. The marginal effect of expected income on annual saving levels is estimated from cross-group variation in average income levels across survey years. If unobserved heterogeneity were an important source of bias in our respondent-level regressions, the group-level regressions ought to yield substantially larger marginal propensities to save out of expected income and smaller marginal effects from unexpected income.

However, as column (1) of Table 7 documents, group-level analysis does not provide such evidence. In fact, the estimated marginal propensity to save out of expected income from the group-level analysis, 0.29, is virtually identical to our baseline parameter estimate shown in Table 6, column 1 (0.27). Furthermore, the group-level regression yields a larger point estimate for the marginal propensity to save out of unexpected income shocks (0.71) than the baseline parameter from the respondent-level regression (0.49 from Table 6, column 1), thereby reinforcing our previous interpretation.

Since we cannot rely on theory to guide our group definitions, we carry out the aggregate-level regression analysis using several alternatives, the results of which appear in the different columns of Table 7. Results based on these alternative group definitions are precisely the opposite from what we would expect to find if unobserved heterogeneity generates biased regression coefficients at the respondent level. Group-level estimates of α_2 exceed those based on the respondent-level analysis, while estimates of α_1 are stable to changes in methodology.

On the basis of the group-level regressions, we dismiss the hypothesis that differences in attitudes toward work and saving at the individual level (lazy, myopic non-savers vs. hard-working, prudent savers) drive our respondent-level regression results. Rather, we conclude that small marginal propensities to save out of expected income and large propensities to save out of unexpected (shocked)

[28] Here, unexplained income simply equals the difference between average annual income among all workers in the group this year and estimated expected and unexpected incomes.

income characterize important behavioral patterns guiding saving decisions among working-class American families a hundred years ago.

V. A Historical Survey of the Institutional Environment to Accomodate Consumption Smoothing among American Workers

Many contemporary financial institutions which allow consumers to postpone payment or to borrow against future income were not much in evidence at the end of the nineteenth century. It seems obvious to state that there were no credit cards, no installment plans, and no home equity lines of credit available in the U.S. during the period under study. Nonetheless, we complete our research by including an inquiry into the institutional environment in which working-class families were making saving and other decisions in the face of risky income prospects.[29]

First, a note about household saving. The primary empirical finding of this paper is that households saved portions of their own income to smooth consumption in the face of uncertain employment prospects. The marginal propensity to save out of transitory income is found to be quite large; our previous work suggests a large average propensity to save out of overall income by working-class households at the turn of the century -- about 7.8 percent of total family income (James, Palumbo and Thomas, 1996). Much of that saving is likely to taken place outside of formal institutions (such as bank accounts), in the shape of cash held in the home, especially for smaller amounts. Thus, much of the savings used to smooth consumption was internal and informal -- spreading income beyond the pay period by adding to, and later dipping into, a domestic fund. Recall that the opportunity cost of holding money during this period was quite low. The thirty years from 1866 to 1896 generally were times of deflation, so the real return to cash balances was positive. Moreover, in a buffer stock model calibrated to developing economies modest cash reserves are optimal and, thus, tend not to be held for very long periods before they are needed to smooth expenditures after a negative income shock (Deaton, 1992, chapter 6; or Deaton, 1990). Finally, liquidity would have been a very important characteristic for buffer stock savings and, as described below, cash might have looked very attractive relative to other financial instruments available at this time.

Financial markets were not entirely absent from assisting with the consumption-smoothing motivations of working-class families. The postbellum period exhibited rapid spread of mutual savings banks. Whereas fewer than 7 percent of American households were bank depositors in 1850, by 1870

---

[29]See also Rotella and Alter (1993, 112), who note that “we know very little ... about the extent to which families in the nineteenth century could and did participate in capital markets.”

the proportion had risen to 21.5 percent, and by 1900 to 38.2 percent. The average deposit also rose; in 1900 prices, from $240.98 in 1870 to $401.09 in 1900. We cannot be certain how far working class households participated in the diffusion of bank accounts; but we should note that fewer than 10 percent of the Michigan furniture workers surveyed in 1889 declared that they had “money at interest or in the bank.”[30] It is possible that this figure underestimates the true proportion. A smaller survey of New Hampshire workers undertaken during 1894 identified 48 percent of workers as having savings accounts, a proportion similar to that produced from a survey of Maine workers in the same year.

In a later nationwide investigation (from 1902) of 1,347 American families who generated a surplus of incomes over expenditures (and reported their behavior), about half (50.3 percent) placed their excess money in the bank (U.S. Commissioner of Labor 1904). According to the same source, 36.5 percent of families who saved kept the money at hand; 7.8 percent invested in either real estate or deposited funds with a building and loan; 4.5 percent paid off existing debts; and only 3 households invested in the stock market.[31] At least a third of the families in this survey owned bank accounts.

When added to the spread of building and loan societies in the late nineteenth century, it becomes clear that financial institutions increasingly accommodated working-class savers. At the same time, such formal arrangements were clearly not the major mechanism for smoothing income fluctuations. One piece of evidence that supports such an interpretation was the common requirement that considerable notice be given before monies could be withdrawn from a savings bank--indicating that deposits in formal institutions were more likely to be used for planned purchases, rather than unanticipated shortfalls in income.[32]

What other mechanisms were open to working-class households to finance shortfalls of income? Besides spending down past savings, some families might have tried to borrow against future earnings. The results of the federal household survey in 1902 reveal evidence on this practice. Of 357

---

[30]However, a similar question asked of Michigan copper miners in 1888 elicited 179 positive responses (out of 2497 workers), while 952 refused to answer. If we treat the refusals as account holders who are loathe to announce the amount of their savings (for fear of taxation, perhaps), the proportion of account holders is 45.2% of those surveyed; if, however, they are treated as “don't knows,” the proportion of (positive) respondents falls to 14.8%.

[31]There were 2,567 households altogether in this survey who reported the balance of incomes and expenditures. A further 133 households who ran surpluses did not disclose their savings activities.

[32]See the evidence of George V. Cresson, President of the Manufacturers' Club of Philadelphia, to the Industrial Commission, December 18, 1900: “You cannot draw the money out [of a savings bank] except after so many weeks' notice” (Industrial Commission, XIV, 271).

families who reported a deficit (expenditure in excess of income) and also reported the source of their finance, 45.4 percent obtained credit, 26.3 percent depleted their stock of savings (whether held in banks or at home), 3.6 percent borrowed money, and the remainder mortgaged or sold property or real estate.[33] It might seem, on this basis at least, that borrowing against future income was a more common route to smoothing consumption than depending on past precautionary or buffer-stock savings. However, in the absence of detailed statistics on the amount of the deficits so financed, such evidence is not conclusive. It may well have been the case that small deficits were financed by running up credit with a local merchant, shopkeeper or landlord, whereas larger deficits were financed out of accumulated assets (liquid or semi-liquid).[34] Informal credit markets, borrowing from family or friends, may have been useful in response to short-term idiosyncratic shocks, such as illnesses or minor accidents, but most likely could not have been relied upon to smooth consumption in the face of large aggregate adverse shocks, such as widespread unemployment, when such requests may well have been overwhelming. Our database, unfortunately, includes no direct information about gift-giving or receiving upon which to draw firm conclusions in this respect.

On the other hand, pawnbroking represents an activity about which more is known. Through a pawnbroker families could liquidate durable goods or mortgage assets to finance current expenditures. The most common currency of the pawnbroker was jewelry (including watches, rings, etc.), with clothing not far behind (other pawned assets included musical instruments and firearms). The rates charged for borrowing against such assets was considerable-- upwards of 1.5 percent per month[35]-- but in the absence of other sources of credit, it was frequently employed, albeit not as frequently as in Europe.[36] The average loan was small--$4.14 in mid-1897-- although the recorded loans in one inquiry included "$10,000 worth of railroad stock" (Patterson 1899, 274). Most loans

---

[33]Note that 150 other households did not declare the means by which their deficit was financed.

[34]Using data from the 1918/1919 Consumer Purchases Survey, Martha Olney (1998) finds that 25.0 percent of white families and 21.7 percent of black families sampled used merchant credit. However she notes that these figures represented just families with credit balances outstanding at the end of the year and hence cannot be used to determine what share of spending was cash versus credit financed.

[35]Patterson's study of pawnbroking revealed annualized interest rates ranging for the loan of $1 for one month from 24% in Kansas City to 300% in New Orleans; for loans of $105 for 4 months, the rates ranged from 24% in Kansas City to 120% in Pittsburgh and Providence (1899, 270).

[36]The average number of loans per capita was between 5 and 6 in Britain before 1914, compared to 0.062 articles per person in Cleveland and 0.0923 per person in Boston in 1897/8.

were short-term--legal arrangements established maximum loan periods at anything from 4 to 12 months, while custom in most cities limited loans to 30 days, although they could be rolled over into a new contract. Much activity was promoted by repeat customers -- those who pawned an item on Monday, to retrieve it on Saturday evening in anticipation of celebration (secular or religious), before pawning it again on Monday. One should not, however, overstate the short-term focus of pawning. The average length of a loan issued by the Collateral Loan Company of Boston, a major pawn agency, was 4.5 months in 1874 and over 3 months in 1897. Although we have only limited information on the motives of pawning, the evidence from Britain suggests that it is most plausible to view it as a response to crises of liquidity, whether originating from unemployment, sickness, or accident (Johnson 1985, 174 ff.). For many working class households, especially those unable to afford regular investment in a mutual savings or building and loan account, physical assets such as jewelry and quality clothing offered both service flows from luxury consumption and a precaution against ill-favor in the future.

New institutions developed in the late nineteenth century in response to the perceived greater risks associated with industrialization. In particular, the development of trade unions as mutual insurance societies proliferated, especially after 1880. The report of the New York Bureau of Statistics of Labor in 1895 identified 909 active labor organizations in the state in 1894 (including both local affiliates of national unions and small independent unions), of which only 141 had been in existence before 1880.[37] None of the national unions had any benefit scheme in place in 1880. By 1894, 78 per cent of the New York labor organizations reported some benefit programs. Similarly, the 1896 report of the Michigan Bureau of Labor Statistics identified 237 labor unions with benefit schemes (of which 21 gave out-of-work benefits, 73 gave weekly sick benefits, 93 gave burial benefits, 58 had life insurance schemes, and 107 gave strike insurance).

The spread of these collective insurance agencies should not be exaggerated. The Michigan organizations canvassed in 1896 employed less than 8 percent of the Michigan non-farm labor force in 1896. To this total ,on the other hand, should be added the non-labor organizations that provided insurance against accident, sickness and other hazards. Thus, the Michigan bureau's report on copper miners in 1888 reported that, "many of the men are members of fraternal organizations... There are no labor unions of any kind among the men, as it is understood that the mining companies are opposed to labor organizing." Our sample of working men indicates that some 27 percent belonged to some kind of benefit society (some workers belonged to more than one) offering sickness, accident, life or

[37]Similar statistics exist for Illinois and Indiana. See Bemis (1899), p. 362.

burial insurance. Clearly, coverage against death does not constitute precautionary saving in any normal meaning of the phrase, so that we should be careful not to ascribe too much protection from risk to the average worker from this figure. Indeed, the proportion of workers carrying accident insurance among Kansas workers in 1895/6 was only 5.5 per cent, even though 26.4 percent belonged to a benefit society of some kind. Moreover, almost no benefit societies (and very few labor unions) insured workers against unemployment, the major source of instability in earnings. Palumbo (1998) finds for 1890 Maine workers no evidence that membership in benevolent societies affected saving behavior, though labor union dues appear to have offset some personal saving.

Overall, it seems plausible to argue that personal saving in the form of cash holdings and other informal arrangements likely represented the dominant instruments used to smooth consumption against unpredictable shocks to earnings. Borrowing from friends and relatives, running up store credit, letting rents fall into arrears, and, most notably, dipping into accumulated savings at home (under the mattress, or in the jar on the mantelshelf) may well have the most important means to maintain consumption in the wake of sickness or unemployment, especially for the most vulnerable. The lower paid, unable to build up a nest egg of cash or having already exhausted it, probably were the most likely customers for the pawnbroker -- whether on a regular or an occasional basis -- and may also have been recipients of such limited private charity as was available in this period to support “the less well-off.” Those with more skills and higher (and perhaps more regular) incomes may have been able to take advantage of more formal institutions, via their membership of benefit societies and their maintenance of mutual savings bank accounts. Note however that even for these households, it would have been necessary to utilize informal agencies to smooth consumption, especially given that most benefit societies required a minimum period before benefits could be paid, and that savings banks required notice of withdrawals.

VI. Summary and Conclusions

In this paper we examine the saving behavior of American workers around the turn of the century using a pooled set of independent cross-section workers’ surveys. In the spirit of Paxson (1992), we estimate unexpected income directly rather than treating it as a residual, using information on the variability of days lost to estimate shocks to time spent out of work (for reasons of illness and accident, as well as unemployment). These types of shocks would have had minimal direct effects on consumption, but would rather have acted on consumption through unexpected changes in income. The shocks to income resulting from unexpected days lost thus serve as an explicit measure of unanticipated income realizations. We use this information to estimate marginal propensities to save out of expected and unexpected incomes in a pooled cross-section regression. We show that American workers around the turn of the century, a time before formal government social insurance programs and before most workers had access to formal credit markets, were still on average able to smooth their consumption relative to income. Our econometric evidence consistently shows much larger estimated marginal propensities to save out of unexpected income than out of expected income, in alternative specifications and also allowing for possible unobserved heterogeneity.

Consistent with several other empirical analyses of family saving behavior by Flavin (1991), Paxson (1992), Alessie and Lusardi (1997) and Pistaferri (1998), our evidence rejects the strict permanent income or certainty equivalence (CEQ) model (Deaton 1992; Browning and Lusardi 1996) in which the marginal propensity to save out of expected income is zero and that out of unexpected income shocks is one. Our estimated saving propensities out of predictable income for late-nineteenth/early-twentieth century American working-class families, ranging between .17 and .34, are quite similar to Paxson's estimates (1992), which ranged between .25 and .28 based on comparable saving definitions, for Thai rice farmers. On the other hand, Paxson's estimated marginal propensity to save out of unexpected income, .74 to .75, are rather larger than our estimates, .48 to .58. Furthermore, our estimated parameters for predictable income movements are very similar to estimates of the change in saving due to predictable changes in income (.18 to .25) from the excessive sensitivity of consumption studies done by Flavin (1991) for American families in the late 1960's, and by Alessie and Lusardi (1997) for Dutch families in the mid 1980's. Pistaferri's (1998) analysis of the saving decisions observed among contemporary Italian families yields marginal propensities to save out of permanent movements in income around 0.16 and out of transitory income shocks between about 0.50 and 1.20. Thus, despite differing specific empirical approaches and microdata from quite different economies, all five of these papers support the conclusion that families use saving to smooth consumption expenditures at a relatively high frequency (such as over the business cycle), but also that saving responds more to expected income movements than it "should", where reference is made specifically to behavior from the permanent income hypothesis.

Further work is needed, however, before stronger conclusions about household saving behavior can be reached. For example, none of the papers cited above use data or methods suitable for examining the ability of families to use saving to smooth consumption expenditures over longer time horizons (such as the overall life-cycle) or at frequencies higher than a few years (such as in response to seasonal volatility). The results of all these empirical papers that reject the CEQ model suggest in turn behavior consistent with more modern treatments of household intertemporal optimization which include explicit liquidity constraints (Deaton 1991), precautionary saving (see Hubbard, Skinner, and

Zeldes 1994), or buffer-stock saving (see Carroll 1997). As surveyed by Deaton (1992) and Browning and Lusardi (1996), these models generally imply marginal propensities to save out of expected income that are greater than zero, and marginal propensities to save out of unexpected income shocks that are less than one. Without data on accumulated net worth, or strong a priori information about the income generating process at the family level, it is difficult to distinguish between these three theories in any particular case (see Deaton 1992; Browning and Lusardi 1996). We leave a more precise investigation into "structural determinants" of saving at the of turn-of-the-century to future research.

From this perspective of contemporary theory, our finding that estimated marginal propensities to save out of expected income movements and unexpected income shocks (taken from the existing literature) seem invariant to the database under study presents a new challenge to researchers. As suggested above, current models of precautionary saving, buffer stock saving, or liquidity constrained behaviors all imply that optimal saving policies should depend on the specific characteristics of the family income process. Thus, if income processes, plausibly, vary across stages of economic development and because of different social insurance policies implemented by different governments, then the marginal propensities to save estimated by others and by us should not be as similar as they appear to be. The current set of similar saving and expenditure regressions generated so far in this branch of the saving literature suggests consideration to these particular empirical patterns in future research.

Finally, the results on savings behavior presented here also have more general implications for our understanding about the motivations for saving among the working-class in the U.S. during the late nineteenth century. In none of three dozen workers' surveys did a majority of households report positive saving during the previous year (James, Palumbo, and Thomas 1996). On the basis of such cross-sectional evidence, Carter and Sutch (1996), for example, assert that there were two types of families that populated the working-class at this time — a minority who saved a large portion of their annual incomes consistently and a majority who never saved at all. However, through the lens of contemporary economic theory based on intertemporal optimization in the face of uncertain earnings, there is no reason to assume such bifurcation in underlying saving motives or tendencies at the micro level. In fact, recent analysis reveals that a single model of intertemporal optimization applicable to all families is quite capable of generating the observed heterogeneity in saving outcomes (frequencies and levels) documented in our cross-section survey data. Thus, Deaton's simulations based on buffer-stock models in the face of liquidity constraints and income uncertainty (1991; 1992) readily generate patterns of frequent annual dissaving and infrequent, but sometimes substantial, annual saving by rational forward-looking families. Indeed, the ability of modern theories based on intertemporal optimization in the face of income uncertainty to generate heterogeneity in savings and wealth accumulation among families might well be the most important implication (Hubbard, Skinner and Zeldes, 1995; Browning and Lusardi, 1996; Carroll, 1997). In the absence of contrary evidence, one might well hypothesize a common behavioral model for all working-class families, rather than assuming that heterogeneous outcomes were generated simply by wide differences in tastes or other underlying motives to save and spend.

References

Alessie, Rob and Annamaria Lusardi (1997), "Saving and Income Smoothing: Evidence from Panel Data," European Economic Review, 41, 1251-79.

Bemis, Edward W. (1899), "Benefit Features of American Trade Unions," Bulletin of the Department of Labor, vol. IV (May), 361-400.

Browning, Martin and Annamaria Lusardi (1996), "Household Saving: Micro Theories and Micro Facts," Journal of Economic Literature, 34 (December), 1797-1857.

Carroll, Christopher D. (1997), "Buffer-Stock Saving and the Life-Cycle/Permanent Income Hypothesis," Quarterly Journal of Economics, 112 (February), 1-55.

Carter, Susan and Richard Sutch (1996), "Myth of the Industrial Scrap Heap: A Revisionist View of Turn-of-the-Century American Retirement," Journal of Economic History, 56 (March), 5-38.

Deaton, Angus (1992), Understanding Consumption. Oxford: Oxford University Press.

Deaton, Angus (1991). "Saving and Liquidity Constraints," Econometrica, 59 (September), 1221-48.

Deaton, Angus (1990). "Saving in Developing Countries: Theory and Review," in Proceedings of the World Bank Annual Conference on Development Economics, 1989. Washington: World Bank. Pp. 61-96.

Flavin, Marjorie (1991), "The Joint Consumption/Assets Demand Decision: A Case Study in Robust Estimation," NBER Working Paper No. 3802.

Hubbard, R. Glenn, Jonathan Skinner, and Stephen Zeldes (1994), "The Importance of Precautionary Motives in Explaining Individual and Aggregate Saving," Carnegie-Rochester Conference Series on Public Policy, 40 (June), 59-126.

Industrial Commission (1901), Report of the Industrial Commission on the Relations and Conditions of Capital and Labor. Washington, DC: Government Printing Office.

James, John A. (1993). "Changes in Economic Instability in 19th-Century America," American Economic Review, 83 (September), 710-731.

James, John A., Michael Palumbo, and Mark Thomas (1996). "Have Working-class Americans Always Been Low Savers? Savings and Accumulation before the Advent of Social Insurance: The United States, 1885-1910," unpublished manuscript.

James, John A., Michael G. Palumbo and Mark Thomas (1999). "What's the Frequency? Consumption Smoothing at High- and Medium-Frequencies among American Workers around the Turn of the 20th Century," unpublished manuscript.

James, John A. and Mark Thomas (1996), "The Golden Age: How Labor Markets Worked One Hundred Years Ago," unpublished manuscript.

Johnson, Paul (1985), Saving and Spending: the Working-class Economy in Britain, 1870-1914. Oxford: Clarendon Press.

Kantor, Shawn and Price Fishback (1996), "Precautionary Saving, Insurance, and the Origins of Workers' Compensation," Journal of Political Economy, 104, 419-42.

Keyssar, Alexander (1986). Out of Work. Cambridge: Cambridge University Press.

Olney, Martha L. (1998), "When Your Word is Not Enough: Race, Collateral, and Household Credit," Journal of Economic History, 58 (June).

Palumbo, Michael G. (1998), "Estimating the Effects of Earnings Uncertainty on Families' Saving and Insurance Decisions," unpublished manuscript.

Patterson, W. R. (1899), "Pawnbroking in Europe and the United States," Bulletin of the Department of Labor, vol. IV (March), 173-310.

Paxson, Christina H. (1992). "Using Weather Variability to Estimate the Response of Savings to Transitory Income in Thailand," American Economic Review, 82 (March), 15-33.

Pistaferri, Luigi. (1998), "Superior Information, Income Shocks and the Permanent Income Hypothesis," mimeo, University College London.

Romer, Christina (1986), "Spurious Volatility in Historical Unemployment Data," Journal of Political Economy, 94, 1-37.

Rotella, Elyce and George Alter (1993), "Working Class Debt in the late Nineteenth Century United States," Journal of Family History, 18, 111-134.

Townsend, Robert M. (1995), "Consumption Insurance: An Evaluation of Risk-Bearing Systems in Low-Income Economies," Journal of Economic Perspectives, 9 (Summer), 83-102.

U.S. Commissioner of Labor (1904), Eighteenth Annual Report, 1903. Washington, DC: Government Printing Office.

Weir, David (1992), "A Century of U.S. Unemployment, 1890-1990: Revised Estimates and Evidence for Stabilization." In R. Ransom, ed., Research in Economic History, Vol. 14, Greenwich, Ct: JAI Press. Pp. 301-346.

Table 1
Summary Statistics for Annual Income and Expenditure
Based on Reported Saving, by Age, Skill and Industry Groups

	(1) No. of Obs.	(2) Expenditure	(3) Income	(4) Var(C)/Var(Y)
A. Total Sample	26,112	442.5	500.3	0.76
Nonsavers	15,377	436.8	436.8	
Savers	10,735	450.7	590.2	
B. by Age Group				
15-25 Years	8661	346.3	382.8	0.79
Nonsavers	5712	338.7	338.3	
Savers	2949	360.9	468.3	
26-40 Years	12258	489.3	558.8	0.77
Nonsavers	6524	498.9	498.9	
Savers	5734	478.5	627.1	
41-55 Years	4377	504.2	569.5	0.73
Nonsavers	2590	497.4	497.4	
Savers	1787	513.9	673.8	
56-75 Years	816	430.7	496.5	0.74
Nonsavers	478	432.5	432.5	
Savers	338	428.1	587.0	
C. by Skill Class				
Unskilled	3698	350.6	380.1	0.90
Nonsavers	2512	356.9	356.9	
Savers	1186	337.1	429.0	
Semiskilled	8820	420.3	469.4	0.79
Nonsavers	5474	413.5	413.5	
Savers	3346	431.3	560.8	

Skilled	9693	503.4	583.3	0.75
Nonsavers	4869	504.3	504.3	
Savers	4824	502.4	663.0	
White Collar	3901	428.9	477.9	0.80
Nonsavers	2449	436.3	436.3	
Savers	1452	416.4	548.0	
D. by Industry				
Manufacturing	15391	409..6	467.6	0.73
Nonsavers	8600	393.8	393.8	
Savers	6791	429.5	561.1	
Mining	935	393.9	432.4	0.85
Nonsavers	624	403.4	403.4	
Savers	311	374.8	490.5	
Transportation	8616	488.5	545.4	0.75
Nonsavers	5457	487.5	487.5	
Savers	3159	490.4	645.4	
Construction	777	569.7	643.0	0.90
Nonsavers	409	608.2	608.2	
Savers	368	526.9	681.7	
Trade & Services	393	589.4	670.0	0.93
Nonsavers	214	638.8	638.8	
Savers	179	530.4	707.2	

Table 2
Summary Statistics for Annual Income and Expenditure
Based on Calculated Saving, by Age, Skill and Industry Groups

	(1) No. of Obs.	(2) Expenditure	(3) Income	(4) Var(C)/Var(Y)
A. Total Sample	11946	457.9	556.7	0.65
Nonsavers	4564	478.4	456.7	
Savers	7382	445.2	618.6	
B. by Age Group				
15-25 Years	1583	386.9	476.8	0.68
Nonsavers	602	412.3	397.0	
Savers	981	369.9	525.8	
26-40 Years	6651	461.5	561.9	0.66
Nonsavers	2438	482.1	462.3	
Savers	4213	449.6	619.5	
41-55 Years	3053	493.5	594.9	0.62
Nonsavers	1235	511.5	483.5	
Savers	1818	481.3	670.6	
56-75 Years	659	429.3	520.5	0.64
Nonsavers	289	443.8	418.6	
Savers	370	418.0	600.0	
C. by Skill Class				
Unskilled	1634	362.7	409.2	0.71
Nonsavers	814	373.5	352.3	
Savers	820	352.0	465.6	
Semiskilled	4288	434.1	515.4	0.71
Nonsavers	1835	455.9	433.9	
Savers	2453	417.7	576.1	
Skilled	5494	505.1	626.6	0.67

Nonsavers	1773	542.2	521.1	
Savers	3721	487.5	676.8	
White Collar	530	454.8	622.8	0.68
Nonsavers	142	574.7	545.1	
Savers	388	410.8	651.2	
D. by Industry				
Manufacturing	8488	445.2	542.5	0.64
Nonsavers	3265	458.3	443.7	
Savers	5223	437.1	604.2	
Mining	484	427.7	496.0	0.66
Nonsavers	199	441.1	381.7	
Savers	285	418.4	575.9	
Transportation	870	574.5	692.1	0.70
Nonsavers	332	605.5	559.4	
Savers	538	555.4	774.0	
Construction	1425	474.1	559.7	0.70
Nonsavers	565	507.1	476.0	
Savers	860	452.4	614.8	
Trade & Services	679	454.4	598.6	0.64
Nonsavers	203	552.2	516.1	
Savers	476	412.7	633.7	

Table 3
Summary of Predicted and Unexpected Workdays Lost during the Previous Year, by Age, Skill and Industry Group

	(1) Average Workdays Lost	(2) Interquartile Range for Unexpected Workdays Lost	(3)
		25%	75%
A. Total Sample	36.5	-28.9	42.3
(N=32,150)			
B. by Age Group			
15-25 Years	38.0	-31.8	16.4
26-40 Years	34.2	-25.7	14.5
41-55 Years	38.1	-26.4	16.7
56-75 Years	48.4	-34.9	22.1
C. by Skill Class			
Unskilled	42.2	-32.7	20.0
Semi-skilled	37.7	-29.6	16.4
Skilled	36.3	-25.7	15.4
White Collar	27.7	-28.9	5.6
D. by Industry			
Manufacturing	36.6	-29.6	16.4
Mining	60.6	-47.6	30.4
Transportation	26.2	-26.9	4.9
Construction	81.0	-44.9	34.5
Trade & Services	22.2	-20.0	-1.6

Table 4
Summary of Predicted and Unexpected Annual Income,
by Age, Skill and Industry Group

	(1) Average Family Income	(2) Interquartile Range for Unexpected Annual Family Income	(3)
		25%	75%
A. Total Sample	506.8	-28.0	49.7
(N=32,150)			
B. by Age Group			
15-25 Years	390.0	-28.2	54.7
26-40 Years	553.7	-25.0	44.2
41-55 Years	573.7	-28.7	45.5
56-75 Years	505.3	-38.0	60.1
C. by Skill Class			
Unskilled	387.2	-34.5	56.3
Semi-skilled	479.1	-28.2	50.9
Skilled	586.1	-26.5	44.2
White Collar	484.5	-9.6	49.7
D. by Industry			
Manufacturing	478.6	-28.2	50.9
Mining	443.6	-52.4	81.9
Transportation	551.1	-8.4	46.3
Construction	586.6	-59.4	77.2
Trade & Services	617.2	2.8	34.1

Table 5
Estimates from Tobit Equation based on Reported Annual Saving
(bootstrapped standard errors in parentheses)

Explanatory Variable	Alternative Specifications			
	(1)	(2)	(3)	(4)
Predicted Income, $\hat{Y}^P$	0.2935 (0.0111)	0.1738 (0.0267)	0.3422 (0.0116)	0.2720 (0.0130)
Unexpected Income, $\hat{Y}^T$	0.5636 (0.0215)	0.5607 (0.0497)	0.5770 (0.0215)	0.5627 (0.0233)
Unexplained Income, $\hat{Y}^U$	0.4298 (0.0094)	0.3570 (0.0191)	0.4368 (0.0094)	0.4263 (0.0094)
Age	---	---	---	3.252 (0.695)
Age-Squared	---	---	---	-0.0418 (0.0091)
Intercept	-184.9 (6.09)	-138.8 (15.67)	-205.1 (6.38)	-230.5 (11.59)

Notes:

(1) contains baseline regression results based on all 26,114 reported saving observations.
(2) contains results based on sample of 26,114 in which all observations have been reweighted such that their sample contributes 10,000 members to the regression. This treatment has the effect of reducing the impact of Michigan surveys on the regression results.
(3) contains results based on a sample of 25,525 observations from Kansas and Michigan surveys only.
4) contains results from a specification based on all 26,114 observations of reported saving with "age" and "age-squared" terms included in the regression equation.

Table 6
Estimates from OLS Equations based on Calculated Annual Saving
(bootstrapped standard errors in parentheses)

	Alternative Specifications			
Explanatory Variable	(1)	(2)	(3)	(4)
Predicted Income, $\hat{Y}^P$	0.2726 (0.0169)	0.2393 (0.0241)	0.2234 (0.0206)	0.3014 (0.0170)
Unexpected Income, $\hat{Y}^T$	0.4933 (0.0245)	0.4885 (0.0387)	0.4771 (0.0293)	0.4931 (0.0242)
Unexplained Income, $\hat{Y}^U$	0.4551 (0.0110)	0.4809 (0.0206)	0.4343 (0.0128)	0.4557 (0.0111)
Age	---	---	---	-5.019 (0.7700)
Age-Squared	---	---	---	0.054 (0.0094)
Intercept	-54.8 (9.20)	-32.6 (13.43)	-28.2 (11.18)	35.0 (17.85)
Adjusted R^2	0.35	0.41	0.30	0.35

Notes:

(1) contains baseline regression results based on all 11,948 observations on calculated saving (reported income minus reported total expenditures).
(2) contains results based on sample of 11,948 in which all observations have been reweighted such that their sample contributes 10,000 members to the regression. This treatment has the effect of reducing the impact of Kansas surveys on the regression results.
(3) contains results based on a sample of 9,306 observations from Kansas and Michigan surveys only.
(4) contains results from a specification based on all 11,948 observations of reported saving with "age" and "age-squared" terms included in the regression equation.

Table 7
OLS Estimates Using Grouped Data on Calculated Saving
(standard errors in parentheses)

Explanatory Variable	Alternative Group Definitions (1)	(2)	(3)	(4)	(5)	(6)
Predicted Income, $\hat{Y}^P$	0.29 (0.073)	0.23 (0.10)	0.27 (0.076)	0.29 (0.14)	0.35 (0.14)	0.43 (0.22)
Unexpected Income, $\hat{Y}^T$	0.71 (0.25)	0.96 (0.24)	1.05 (0.33)	1.06 (0.29)	1.24 (0.16)	0.73 (0.11)
Unexplained Income, $\hat{Y}^U$	0.30 (0.065)	0.38 (0.036)	0.15 (0.10)	0.14 (0.088)	0.33 (0.045)	0.50 (0.055)
Intercept	-62.03 (43.90)	-29.88 (58.79)	-52.25 (45.56)	-71.25 (81.74)	-92.04 (77.02)	-140.83 (52.18)
R^2	0.36	0.43	0.35	0.28	0.43	0.78
No. of Cells	76	79	50	51	111	55
No. of Respondents	10,678	9,877	3,520	3,100	10,887	5,410

Notes:

(1) Regression based on skill-by-industry groupings.
(2) Regression based on age-by-industry groupings.
(3) Regression based on state-by-skill-by-industry groupings.
(4) Regression based on state-by-age-by-industry groupings.
(5) Regression based on birth cohort groupings.
(6) Regression based on birth cohort-by-industry groupings.

Appendix A: Detailed Analysis of Distributions of Reported and Calculated Saving

Left-censoring in reported savings ought to produce a higher average level of savings than that based on calculated saving, which takes some negative values. As it happens, the mean values of the variables are reversed relative to expectations -- average reported savings equals \$57.74 per year, compared to \$98.85 for calculated saving. Some of this gap can be traced to to differing compositions of the two samples of families, in terms of their ages and income levels. The subsample for whom saving is reported directly in the surveys is both younger (about 32 versus 37 years, on average) and poorer (annual income averages \$500 versus \$557) than the subsample for which we calculate saving as the discrepancy between reported income and expenditures. Correcting for compositional differences reduces the gap between the two groups considerably -- we calculate that average calculated saving would only have equalled \$79.72 if its sample displayed the age and income composition of the "reported saving sample"-- but the sign on the gap is still counter-intuitive.

The further explanation lies in the tendency for families either to understate their reported saving or to overstate their reported incomes relative to their reported expenditures. This can be deduced by examining reported and calculated saving among the 7,957 families for which both variables are available. Calculated and reported saving are highly correlated in this subsample. A regression of calculated saving on reported saving yields a slope coefficient equal to 0.9930 (R^2 = 0.53), with an intercept of \$27.50, indicating that households understate directly reported saving relative to surplus income over expenditure. This result suggests that households may have taken a question on saving to mean literally, "money put away," rather than accidental differences between income and expenditure. It might also mean that working households were not always able to account for all of their expenditures.

This divergence however turns out to have virtually no consequence for our empirical analysis. Given the strong linear relation between reported and calculated saving, it should not be surprising that none of the econometric results in this paper are sensitive to the choice of dependent variable, reported saving or calculated saving. As we report below, Tobit equations estimated using reported saving yield nearly identical slope coefficients to OLS regressions based on calculated saving, though the estimated intercepts differ between the specifications.

Appendix B: Algebraic Representation of the Micro-level Econometric Model

This appendix presents the algebraic representation of the micro-level econometric saving model described in Sections II and III. The complete model consists of three regression equations--one for predicting workdays lost (and then for estimating unexpected workdays lost) during the previous year; a second for explaining family income as a function of worker characteristics and unexpected workdays lost; finally, a third for estimating the marginal propensities to save out of predictable differences and unexpected shocks to family income. The regression equations and fitted value definitions permit a simple algebraic representation:

(A1)
$$DL_{i(t)} = X^{p}_{i(t)}{}'\,\gamma + \varepsilon^{DL}_{i(t)} \;,$$
$$\widehat{DL}^{T}_{i(t)} = DL_{i(t)} - X^{p}_{i(t)}{}'\,\hat{\gamma} \;,$$

(A2)
$$Y_{i(t)} = X^{p}_{i(t)}{}'\,\beta^{P} + \widehat{DL}^{T}_{i(t)}\,\beta^{T} + \varepsilon^{Y}_{i(t)} \;,$$
$$\hat{Y}^{P}_{i(t)} = X^{p}_{i(t)}{}'\,\hat{\beta}^{P} \,,$$
$$\hat{Y}^{T}_{i(t)} = \widehat{DL}^{T}_{i(t)}\,\hat{\beta}^{T} \,,$$
$$\hat{Y}^{U}_{i(t)} = Y_{i(t)} - \hat{Y}^{P}_{i(t)} - \hat{Y}^{T}_{i(t)} \,,$$

(A3)
$$S_{i(t)} = \alpha_0 + \alpha_1 \hat{Y}^{P}_{i(t)} + \alpha_2 \hat{Y}^{T}_{i(t)} + \alpha_3 \hat{Y}^{U}_{i(t)} + \varepsilon^{S}_{i(t)} \,.$$

$DL_{i(t)}$ stands for the number of workdays lost during the previous year reported by family i which was surveyed during year t. $X^{P}_{i(t)}$ is a vector of interacted indicator variables describing the worker's characteristics. Then $\widehat{DL}^{T}_{i(t)}$ denotes our estimate of the unexpected component of the family's workdays lost realization. Similarly, $Y_{i(t)}$ is the family's reported income realization; $\hat{Y}^{T}_{i(t)}$ denotes our estimate of the unexpected component to income, while $\hat{Y}^{P}_{i(t)}$ is our estimate of the expected part of realized income. Finally, the residual, $\hat{Y}^{U}_{i(t)}$, measures the part of realized income which cannot be attributed either to the expected or to the unexpected component based on the regression equations.

The three equations of the complete model (A1), (A2) and (A3) are estimated sequentially. First, we estimate equation (1) to obtain coefficient estimates to predict workdays lost for each family and, thus, to allow unexpected workdays lost to be calculated as the residual from (1). Second, we estimate regression (2) using the first-stage residual to measure unexpected workdays lost in the income equation. We then use the second-stage income coefficients to estimate expected income, unexpected income and unexplained income (the residual from (2)) to estimate the saving regression equation (3). Bootstrap methods are used to estimate consistently standard errors for the coefficients in the third-stage saving regression.

Table A-1
Survey Composition of the Dataset

	Surveys with Reported Saving		Surveys with Calculated Saving	
State	Year	No. of Obs.	Year	No. of Obs.
Maine			1886.5	62
			1887.5	87
			1888	88
			1890	1011
			1894	504
			1900	102
New Hampshire			1886	42
	1887	92	1887	226
Michigan	1888	715		
	1889	4265	1889	1910
	1890	5920	1890	3801
	1892.5	5028		
	1895	3001		
	1896	3757		
Kansas			1884.5	349
	1885.5	316	1885.5	403
	1886.5	323	1886.5	393
	1895	233	1895	384
	1896	302	1896	426
			1899	724
	1903	617	1903	594
	1904	321	1904	320
	1905.5	329		
	1906.5	396		
Missouri	1890.75	255	1890.75	163
Oklahoma	1908	242	1909	117
West Virginia			1893.3	170
Washington			1900	32
			1902	38