Re-estimating the Gainful Employment Rate of Older Men: the United States, 1870 to 1930.

Gardner, John and Gratton, Brian and Moen, Jon

April 2018

Re-estimating the Gainful Employment Rate of Older Men: the United States, 1870 to 1930

John Gardner
Department of Economics
The University of Mississippi
jrgardne@olemiss.edu

Brian Gratton
The Department of History
Arizona State University

Jon Moen
Department of Economics
The University of Mississippi
jmoen@olemiss.edu

Abstract: Analyses of the economic effects of the introduction of the public pension system on older men in the US have been hamstrung by difficulties generating reliable estimates of historical labor-force participation rates using data from early US censuses that only asked respondents about their occupations and not whether they were actively employed. We extend a unique feature of the 1901 Canadian census, which asked about retirement status as well as occupation, to older men in the 1900 US Census to estimate labor-force participation rates that adjust for misreporting of employment status. Our estimates show that reported rates substantially overestimate labor-force participation among older men. We also show that adjusted rates based on an econometric correction for misclassified limited dependent variables produces are similar to those based on the 1901 Canadian census. Using this technique to extend our adjustment shows that reported rates overstate older men's labor-force participation rates in the 1880, 1910, 1920 and 1930 census, as well as the decline in those rates between 1900 and 1910.

Keywords: Labor-force participation, social security, retirement, measurement error, misclassification.

JEL Codes: J21, H55, J26.

A partially resolved debate exists among economists studying the labor force behavior of older men before public pension systems. It remains unresolved because data on work and retirement before 1940 are not comparable to that collected in U.S. censuses since 1940. The earlier census data also contains occupational misreporting errors that obscure trends in labor force behavior of older men. The debate's significance lies in its bearing on theoretical assumptions about such behavior in the absence of exogenous social provision for retirement and in the practical difficulty of estimating accurately the impact of such provision without sound data for the earlier period. The main goal of this paper is the adjustment of pre-1940 census data, creating labor force participation rates (LFPR) comparable to contemporary LFPR.

In the U.S., the 1940 Census provided the first measurement of LFPR as currently used to assess labor force activity. Before that year, census enumerators collected gainful employment data, asking respondents to state what occupation they had. But the enumerators did not inquire whether the respondents were working or actively seeking work. As discussed below, many older men reported an occupation, even though they no longer practiced it. In the first, uncritical use of gainful employment data from censuses taken in the largely agricultural economy of the 19th century, investigators reported very high levels of work activity, reaching four/fifths of men 65 and over; these studies found steady declines in activity across the late 19th and early 20th centuries, and then abrupt decline after Social Security was instituted in the 1930s, falling from 58% in 1930 to 42% in 1940 (Durand 1948). Two schools of thought dominated social scientific views of the economic circumstances of older men and how these circumstances influenced their labor force behavior. The first and dominant interpretation asserted that farm work sustained work activity in older age, while industrialization reduced the control older men had over employment; the factory initiated a steady decline in their participation and impoverished them (for a review of the early interpretation see Gratton 1986) The second, minority model suggested that the economic growth engendered by industrialization did not necessarily undermine the competitiveness of older men in the labor market (Long 1958), and that the greater income it produced might have financed voluntary withdrawal, or retirement.

The initial interpretations certainly misinterpreted gainful employment as equivalent to LFPR, and economists who helped design the labor force measure knew that the gainful employment measure exaggerated labor force participation (Edwards 1943; Durand 1948). The pre-1940 data could not be directly compared to measurement of points and trends using LFPR. Because gainful employment surely inflated the number of men said to be working, it would, for example, overstate the exaggerate the effect of Social Security on labor force withdrawal as measured in 1940 and beyond.

Efforts have been made to correct measurement. In an early revision, Durand (1948) used a 1930s unemployment survey to adjust the gainful worker estimate of 1930 to a LFPR; using the Durand method, the 58% gainful employment rate for 1930 fell to 54%, sharply reducing the fall in participation after the institution of social security. Durand and others (Bancroft 1958) extended these corrections backward in time, such that, for example, the gainful employment rate in 1900 of 68% fell to a LFPR of 63%. Subsequent efforts to adjust gainful employment in previous censuses became more sophisticated, relying on other evidence. Ransom and Sutch (1986) argued for steep discounting of gainful employment rates in 1880, reducing the rate of nearly 80% to a LFPR of 64%. They argued that questionable recording and tabulation procedures used by the enumerators of the decennial censuses in 1880 and after led to the inclusion as gainful workers of many men who were unemployed, or, as they maintained, retired. By excluding men who reported 6 or more months of unemployment (and a few minor other classes of misinterpreted occupations like capitalist, landlord, etc), Ransom and Sutch produced very low LFPR. Moreover, LFPR did not decline, but stayed relatively steady across the remainder of the industrial period before Social Security and was only marginally affected by mas public provision.

These conclusions challenged two standard assumptions in the economic literature, the first, that well before the full impact of industrialization or the arrival of Social Security, many older men had left the labor force, and, the second, that Social Security had almost no effect on the overall trend in LFPR in the 20th century. Ransom and Sutch interpreted the low 19th century rates as evidence of higher rates of voluntary retirement, though what they measured was unemployment. Countering this account, Moen (1987; 1988) argued that there was no obvious reason to exclude older men reporting 6 or more months of unemployment from the count of workers, and found that, in 1880, labor force participation was about 78%, declining in the familiar monotonic trend after that date. Margo (1993) and Moen (1994) subsequently showed statistically that the unemployed men would have been counted as part of the labor force, although they were more likely than men reporting fewer than 6 months of unemployment eventually to leave it. Lee (1998a, 1998b) confirmed their results for 1900 and 1910.

Even as economists debated how to measure LFPR, various studies attempted to identify factors that increased or lowered work activity among older men before Social Security. Moen (1994), Costa (1995) and Carter and Sutch (1996) found that, contrary to standard social science theory about agricultural economies, decline in LFPR between 1900 and 1920 was often the result of farmers retiring. This is a position strongly supported by Haber and Gratton's (1994) findings on the elderly household in rural areas of the U.S. and by Dillon, Gratton and Moen (2010) on older Canadian men in 1901. Although these studies found retirement outside the industrial zone in a variety of times and places, Lee (2002) maintained that the American case was a temporary effect of an unusual run-up in the value of farms.

Costa (1995, 1998) reported another source for withdrawal from the labor force, one linked to contemporary findings of pure income effects. She demonstrated that at the turn of the century, U. S. Union Army veterans used the pension to retire, which she notes as an example of a pure income effect. Gratton (1996) argues that by the early twentieth century, working class households could and did accumulate a surplus sufficient to retirement, using their own earnings and those of other members of the household, usually older children. The family-based economy allowed high levels of consumption and significant savings. He estimated that use of the earnings of children allowed aging parents to amass savings sufficient for about 10 years of retirement—more should occasional work be maintained. Moen and Gratton (1999) demonstrated that retirement did occur and in the worst of times. During the 1930s depression, some men were retiring and living on savings. Recent Canadian research agrees with the more positive assessment of older men's economic status, though, the authors do not uniformly conclude that this led to withdrawal from the labor force (Dillon, 2008, Snell, 1996; Montigny, 1997; Baskerville and Sager, 1998).

Some economic historians (as well as most other social scientists) disputed this view, confirming the conventional theory that industrialization left many older workers unemployed, partially employed, or dependent. Lee (2003) found that farmers, professionals, managers, and proprietors were less likely to be lacking work than were craftsmen, salesmen, and operatives. Countering Gratton, he finds that economic well-being of individuals, as measured by consumption expenditure, declined substantially as persons aged. Aged men who were not working tended to be poorer than active workers, and the proportion not working rose sharply after the late nineteenth century.

A new approach to the problem

As census takers in the pre-1940 era were aware, older men often claimed an occupation even though they had long ago laid down their tools. Census instructions in England and Wales in the late nineteenth century urged persons who had retired from their "profession, business or occupation" to state their former occupation alongside the word 'Retired' (Higgs 1982). U.S. Census officials in 1910 were aware that older men would often report the occupation they had previously followed. Instructions to the enumerators point out the following, though there is little evidence they led to more careful enumeration:

> 148. Persons retired or temporarily unemployed--Care should be taken in making the return for persons who on account of old age, permanent invalidism, or otherwise are no longer following an occupation. Such persons may desire to return the occupation formerly followed, which would be incorrect. (1910 U.S. Population Census, Instructions to the enumerators).

We address precisely this issue -- the reporting of occupations no longer followed -- in two ways. First, using a unique data source, the 1901 Canadian Census, we estimate LFPR for the pre-1940 era in the U.S. We first adjust gainful employment data for the 1900 U.S. Census. In subsequent research, we will adjust those of 1870, 1880, 1910, 1920 and 1930 (there is no microdata sample for 1890). In a second approach, we test an econometric correction for misclassified binary dependent variables, and extend the resulting correction to the 1880, 1910, 1920 and 1930 censuses (there is no microdata sample for 1890). All census information used in the project is available through the microdata sample of the 1901 Canadian Census (Sager, Thompson and Trottier 2002) and microdata samples from the IPUMS_U.S.A. project at the University of Minnesota (IPUMS). The 1901 Canadian microdata sample is a 1/20 sample, while our samples of the U.S. IPUMS samples are either 1/100 or 5/100 samples, providing very large representations of the population of men 60 and over.

The Canadian Census of 1901, while remarkably similar to the U.S. Census of 1900, contains one bit of information that can cast new light on older men's labor force activity. In its conventional questions, the Canadian schedule followed the gainful employment approach common to most advanced countries' censuses. However, in addition to asking about the respondent's usual gainful occupation, the Canadian enumerators were required to ask if the respondent had *retired* from that occupation (Dillon, Gratton and Moen 2010). If the answer was yes, the enumerator was to write "r" in the column next to the occupation listed on the manuscript schedule. Thus, it is possible to identify men who were no longer actively seeking work, that is, no longer in the labor force, even though they reported an occupation, eliding the crucial failing in other historical censuses. The enumerators in the corresponding U.S. censuses did not ask this pointed question, despite the known unreliability of older men's reports. The chronological concordance and the otherwise close similarity in the data collected in the two censuses provide a unique opportunity to use the Canadian Census information to adjust estimates of LFPR in the U.S. in 1900. The approach depends on an assumption, one we think well-founded, of appreciable similarities between the two societies in economic, social, demographic, and cultural terms, making it likely that behavior among older men in Canada would be likely among older men in the U.S.

Table 1 shows the total number of men in a set of age groups, 60 and above, in Canada and the U.S., calculated from the 1900 U.S. and 1901 Canadian microdata samples. (U.S. censuses are taken on the decade, while the Canadian censuses occur at the year ending in 1.) It then displays gainful employment, a simple function of the proportion of men who report an occupation. For Canadian men, the percentage who acknowledged retirement can be subtracted, and LFPR estimated by using that correcting factor. From the unadjusted U.S. data, only the gainful employment measure exists. As demonstrated by previous historical studies, in both

Canada and the U.S. most older men reported an occupation; at the turn of the century, about four in five U.S. and Canadian men aged 60 years or more declared an occupation on the census. Nevertheless, factoring in acknowledgement of retirement reshapes Canadian elderly men's work activity, especially among the very old. Among men aged 60 to 64, about 10% record no occupation but an additional 7% "dual report," i.e., state an occupation but also have the letter "r" or the words "retirement" or "rentier" (the term used in French-speaking regions) recorded. For those 70 and over, however, in addition to 29% who report no occupation, another 27% dual report, leaving only 44% in the labor force. It is impossible to identify the status of U.S. men with similar precision. Many U.S. elderly men did report no occupation or gave a non-occupational response, indicating they were no longer in the labor force. But some unknown percentage, like their Canadian peers, gave the name of the occupation they had traditionally practiced, even though they were no longer working. The advantage of the Canadian census was the persistent and systematic inquiry: even if an older man reported an occupation, the census enumerator then asked if he had retired from the occupation. The adjusted rates for Canada provide estimates more consistent and directly comparable to modern LFPR. They also show that adjusting for misreporting of occupations produces downwards revisions at least as large as previous attempts based on reported months of unemployment.

A more detailed look at LFPR at all ages reveals differences across the life course in Canada and the U.S. Figure 1 uses data for individual years of age to contrast the proportions of all men who reported an occupation to U.S. or Canadian enumerators in 1900/01. The results exhibit the pre-Social Security bell-like curve, differing from the modern series both in the much younger entrance of men into work and the lack of a sharp drop-off at official retirement ages. Although the curves have a comforting overall similarity linked to the similarities of the two countries, we do not expect perfect conformity, since Canada and the U.S. had differences in economic, social and cultural characteristics. For example, a much higher percentage of Canadian workers were farmers, an occupation in which the difference between active work and retirement has been seen as particularly ambiguous. The unadjusted series do differ. Canadian men began reporting an occupation on the census slightly later in life than their U.S. counterparts. At age 15, half of U.S. fifteen-year-olds reported an occupation while only 38 percent of Canadian youth did so. This gap continued into the thirties, and, in the late 60s, reversed itself. Older Canadian men were more likely than American men to report an occupation.

Figure 1 then adjusts the Canadian rate for men who reported an occupation but then acknowledged retirement. As expected, there are no differences in the Canadian series until the late 50s, when the removal of retired men profoundly reshapes the view of working life. From the age of 70 on, LFPRs were at least 10% below stated occupational rates. The Canadian series now falls below the American series at all ages past 60. However, the U.S. gainful employment rate surely also exaggerates LFPR, indicating that an adjustment to the U.S. figures is called for. How might the Canadian data be used to make a similar adjustment of U.S. gainful employment measures?

Preliminary research

In an extensive examination of the circumstances of older men in Canada using the 1901 Canadian microdata sample, Dillon, Gratton, and Moen (2010) analyzed the dual reporting of occupation and retirement. The intent was to gauge what factors contributed to a man being more or less likely to dual report. The research proved useful, identifying social and economic characteristics associated with a higher likelihood of reporting an occupation but acknowledging that it was no longer followed. The research also demonstrated that certain factors had little or no effect on that probability. The full logistic regression model from Dillon, Gratton, and Moen (in press) includes age, occupational category (confirming the conventional hypothesis that farmers were the most likely to report an occupation they no longer practiced), urban residence (disconfirming the conventional hypothesis that dual reporting was a rural, traditional phenomenon), regional setting (with men from the most economically advanced region most likely to dual report), other workers in the household and lack of property ownership. Other models revealed effects of literacy, family and household relationships, and also showed that several factors exhibiting cross-tabular correlations, such as ethnicity, had none in multivariate models.

Most of these Canadian variables have exact analogues in U.S. census data. The revealed associations are likely to apply in the U.S. case, indeed in any economy with broadly similar features. The basic logic of the adjustment is to employ the factors shown in the Canadian research to affect the probability of a dual report of occupation and retirement. We apply these adjustments to men in the U.S., looking at progressively older age groups to take into account the known increasing likelihood among older Canadian men to acknowledge retirement while reporting an occupation.

Analytical model

Our first step is to estimate a logistic regression model using the 1901Canadian Census data predicting the LFP rate of Canadian men 65 and older. Other age groups will be separately modeled. The dependent variable is defined as one if an occupation was recorded and no “r” was recorded; it is zero if an occupation was recorded but retirement acknowledged or if no occupation was recorded. We experimented both with models that use only the significant variables found in the Canadian case for dual report (listed above), and with all variables common to or highly similar in Canadian and U.S. censuses. The second step is to estimate the means of the independent variables using the U.S. data.

Equation 1) gives the probability of a Canadian man being in the labor force as predicted by the logistic regression):

1) $P_i = 1/(1 + \exp(-\alpha - \beta_1 x_{i1} - \beta_2 x_{i2} - \ldots - \beta_n x_{in}))$,

which is derived from the logistic regression model

2) $\log(P_i/1-P_i) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_n x_{in}$

We then insert the overall U.S. means estimated from the 1900 U.S. Census for each variable for men of different age groups (e.g. 65 and older) into equation 1) estimated from the Canadian sample. The following age groups will be examined separately: 60-64, 65-69, and 70+. The predicted LFPR can be interpreted as the LFPR U.S. males of this age group would have if their status could be measured as accurately as that of Canadian men. Alternatively, values from individual observations for older U.S. males for the several age groups can be substituted into the 1910 Canadian regression equation, and the resulting predicted values can be averaged by age group.

Table 2 shows first the coefficients of the logistic regression on the Canadian data, indicating the importance of rising age, for example, in making it more likely that retirement would be acknowledged, and that being literate, in contrast, would make it less likely that both an occupation and retirement would be reported. The second column presents the means of the regression variables estimated from the US Census data. In Table 2 the row "Estimated US LFP" presents estimates of LFP for men 65 and older for the US before any attempt to adjust them with the Canadian data. The row “Adjusted US LFP” contains estimates of LFP adjusted with the Canadian Census as described earlier.

As can be seen in the last row of Table 2, the adjustment reduces the gainful employment rate of 71 percent to a rate of 64 percent. The application of the US means for the entire population 65 and over could be considered inappropriate because of very strong differences in demographic, social and economic settings in regions of the US. Perhaps the most important distinction in 1900 is that there is no part of the Canadian population that corresponds closely to the Black population in the US around 1900. While there are other differences—the French-Canadian population stands out—the Black population, subject to strong forms of discrimination like the Jim Crow Laws in the South, as well as racist reception in the prosperous north, remained largely confined to a rural, agricultural, and backward economy, one characterized by sharecropping. Invariably, gainful employment calculations show very high rates for blacks in comparison to whites, even when confined to the farming sector. Although upward mobility toward property ownership was possible and did occur for emancipated blacks (see Higgs, 1980), the level of farm ownership in the South among blacks was the lowest of any group in the United States. This factor alone would suggest that retirement from the labor force on the basis of asset accumulation was least likely in this group. Moreover, sharecropping relied on a family economic effort and older men might continue to contribute to the work effort, however modestly, and they would legitimately contend that they were working in farming.

To account for these demographic and economic differences we re-estimated the means of the independent variables, applying three definitions of comparable US universes. These universes include 1) the US excluding the Cotton South states, 2) including only states that bordered Canada or the Great Lakes, and 3) excluding all Blacks from the estimates. The means and adjusted LFP rates are presented in Table 3. In Table 3 the row “Estimated US LFP” presents estimates of LFP for men 65 and older for the US and the three universes before any attempt to adjust them with the Canadian data. The row "Adjusted US LFP" contains estimates of LFP adjusted with the Canadian Census as described earlier. In the three cases case, the US rates are adjusted downward by 5 to 11 percentage points. This indicates a reasonable range for the number of men who might have been misreporting their labor force status.

The adjustments based on different universes also reflect important features of the US labor force. Removing Blacks from the US means results in the largest downward revision of LFP, about ten percent. This indicates that misreporting was largely a white phenomenon, perhaps because even the possibility of retiring was unlikely for Blacks in 1900. Low income and a strong connection to southern agriculture certainly made retirement a remote possibility at best. Removing the Cotton South states reduces the adjustment to LFP by about seven percentage points, while limiting the US sample to Canadian border states reduces the adjustment by about five points. Both subsamples also have lower unadjusted LFP rates than the US as a whole. The adjustment to US estimates appears to get smaller in states farther away from the South. The US adjustment is also smaller than the direct adjustment made to Canadian LFP rates, which for men 65 and older falls from 76 to 56 percent (see Table 1). A larger agricultural population in Canada might account for some of the difference between the US and Canadian adjustment. The results from the Canadian Census and the adjustment made to the US Census indicates that adjustments made to nearby US Censuses will prove fruitful and are likely to alter our understanding of labor force activity in the later 19th and early 20th centuries.

Age-group-specific LFP adjustments, shown in Table 4, vary as expected with age in each of the universes that we consider.[1] At younger ages, both the likelihood of being out of the labor force and of misreporting labor-force participation status are much lower than at older ages.

[1] To improve the precision of the estimated age-group-specific means, we base this analysis on 5% IPUMS extract of the 1900 US census.

Misreporting of LFP status increases markedly at age 65. Combined with the fact that the populations of men 65 and older are relatively small means that the LFP adjustment is much more severe for men 65 and older than for those 60 and older (compare the adjustment from 71 to 63% for all men age 65 and up reported in Table 3 with the adjustment from 77 to 73% for the population of men aged 60+ reported in Table 4).

Unfortunately, no other Canadian census enumerated using the concept of gainful employment used the direct retirement inquiry, making it impossible to follow a similar comparative procedure. But an alternative approach to re-estimating LFPR from gainful employment data will be employed for all U.S. censuses between 1880 and 1930. Furthermore, these estimates can be compared to those made by other researchers such as Durand, Bancroft, and Ransom and Sutch.

Misclassification Adjustment

The alternative approach uses econometric techniques designed to handle misclassified binary dependent variables and is independent of the first adjustment procedure. Comparing the results from the two procedures will also provide a robustness check for the adjustments made on the 1900 U.S. census data. The dependent variable in a logit (and probit) regression model runs the risk of being misclassified, that is, a one being assigned when in fact a zero should be assigned. In the case of estimating the labor force, the misclassification arises when a man reports an occupation to the enumerator and is assigned a value of one in the regression, indicating being in the labor force, when it he should actually be assigned a zero because he no longer actually follows the stated occupation. While it is possible that some older men might not report an occupation when they actually are following one actively, this seems much less likely than incorrectly reporting an occupation. That is, there are few older men who would not report an occupation and be recorded with a zero when in fact they do have an occupation and should be recorded with a one in a machine-readable data set like the IPUMS samples.

The theory behind misclassified dependent variables is clearly laid out in Hausman, Abrevaya, and Scott-Morton (1998) with an application to job switching. Other applications include self-reporting of student cheating (Caudill and Mixon 2005), and labor market outcomes (Falaris 2009). These studies reveal that misclassification of the dependent variable in a logit or probit model results in inconsistent coefficient estimates when no correction is taken into account.

The evidence presented thus far strongly supports the related conclusions that in the 1900 US Census, some older, retired men were misreported as being gainfully employed and that this misreporting exaggerated labor-force participation rates based on 1900 Census Microdata. To redouble the evidence on the extent and consequences of such misreporting, and to investigate the persistence of such reporting over time during the early 20th century, in this section we estimate a series of misclassification-corrected probit models that account for potential misreporting of employment status (Hausman, Abrevaya and Scott-Morton, 1998).

In the misclassification-corrected model, while *actual* employment status follows a standard probit model, *reported* employment status also depends on the probability α of erroneously reporting gainful employment (i.e., an occupation) conditional on actually being retired (i.e., the probability of a false positive). In this setting, a man may report gainful unemployment either because he is employed (which occurs with probability $\phi(x_i\beta)$, where x_i is a vector of covariates and β is the corresponding vector of coefficients, and ϕ is the standard normal CDF) and reports so faithfully or because he is not employed (with probability $1-\phi(x_i\beta)$) but reports being so with probability α. Thus, the likelihood of reporting employment is $\phi(x_i\beta)+\alpha[1- \phi(x_i\beta)]=\alpha + (1-\alpha)\phi(x_i\beta)$, and similarly for unemployment. The sample log likelihood function is therefore

3) $\log L(X; \beta, \alpha) = N^{-1} \sum_i \{y_i \log [\alpha + (1-\alpha)\phi(x_i\beta)] + (1 - y_i) \log[1 - \alpha - (1 - \alpha)[1 - \phi(x_i\beta)]]\}$

(also see Caudill and Mixon, 2005, for an intuitive derivation). For the sake of simplicity, this formulation of the misclassification-corrected model assumes that the probability α of a false positive does not depend on the covariates.[2] In keeping with intuition given above, it also assumes that men who are gainfully employed always correctly report their employment status (i.e., false negatives occur with probability zero).

To improve the precision of our initial estimates, we use a 5% IPUMS sample of the 1900 Census (Ruggles et al, 2015). The dependent variable is an indicator for (self-reported) labor-force participation status, and we use the same covariates as in the previous regressions (age, white, farm, rural nonfarm, native, head, literate, marital status, and number of children). In order to focus on older men while maintaining a reasonably large sample, we estimate the model using observations on all men aged 60 and up.

Table 4 reports standard and misclassification-corrected probit estimates of labor-force participation status. The misclassification-corrected coefficient estimates have the same signs as their uncorrected counterparts, but larger magnitudes. This pattern reflects the well-known tendency of classification error to attenuate the apparent relationship between variables. The estimated probability of a false positive (that is, reporting an occupation when, in fact, retired) is .22, which concords well with the proportions reported in Table 1 of older men in the 1901 Canadian Census reporting an occupation who also report retirement. Those proportions range from .07 to .27, with an average of .16. The similarity of these numbers to the estimated probability of a false positive among older men in the 1900 US Census suggests that the misclassification-corrected probit does a good job of identifying potential misreporting. At the same time, the higher

[2] This does not, however, mean that difference between reported and corrected labor-force participation rates are independent of the covariates, since the covariates do influence the true probability of employment.

probability of a false positive among American men suggests that demographic differences between the US and Canada do generate differences in the misreporting of employment status. We also use the misclassification-corrected probit estimates to generate labor-force participation rates that adjust for such misreporting. In particular, we combine the covariates with the misclassification-corrected coefficient estimates in order to predict the probability of being a labor-force participant for each individual in the 1900 sample of older American men (the predicted probabilities are given by where are the estimated coefficients from the misclassification-corrected probit). We then average these individual probabilities within age groups to estimate age-group-specific labor-force participation rates.

Table 5 summarizes the estimated labor-force participation rates. For comparison, we also present rates based on simple averages of reported labor-force participation and predicted probabilities from a standard probit model of labor-force participation. Adjusted rates based on participation probabilities derived from the misclassification-corrected probit are substantially lower. Among all men aged 60 and up, the adjusted rate is .71, in comparison to a reported rate of .77. As progressively older age groups are examined, the adjusted labor force participation rate declines steeply from .85 among men aged 60-65 to .52 among men aged 70 and older. The impact of the adjustment increases similarly with age; in the oldest age group, the adjusted rate is fully 10% below the reported rate. The LFP rates, adjusted in this manner, correspond closely to those adjusted using the 1901 Canadian census reported in Tables 3 and 4, strongly suggesting that both methods recover something close to the true rates LFP.

In Table 6, we extend this analysis in two ways. First, because our previous analysis based on the Canadian census showed that excluding blacks from the sample generated the largest difference between reported and adjusted participation rates, we estimate separate models using samples of white and black men. Second, to explore how the misreporting of employment status, and labor force participation rates themselves, changed throughout the early 20th century, we also estimate misclassification-corrected models using samples of the 1880, 1910, 1920, and 1930 Censuses.[4] We note that for these later Census decades, only 1% samples are available through IPUMS, which reduces the precision of our estimates, particularly for the relatively small samples of black men.

As the table shows, using separate 1900 Census samples of white and black men results in a more severe estimated rate of misclassification among white men, and a lower adjusted rate of labor-force participation for white men. In contrast, misreporting had no apparent effect on the estimated rate of labor-force participation for black men in that year, among whom 88% of those 60 and older were active participants. This trend is also similar to the adjustment based on the Canadian census presented in Table 3.

Focusing first on white men, Table 6 shows two clear time trends. First, though rates of labor force participation declined between 1900 and 1930, accounting for misclassification reduces the estimated decrease (from nine to four points when considering all men aged 60 and up). Second, the false positive rate declined substantially, from about 23% in 1900 to only 6% by 1930, possibly reflecting better practices among enumerators in later decades. For black men, both the raw and adjusted rates evince a greater decline in labor-force participation (which started at a much higher rate) than for white men, though the adjusted decline is slightly larger. However, some of the black samples are quite small, and we have less confidence in the estimated probabilities of false positives, which have large standard errors and fluctuate considerably between decades).

Conclusion

---

[4] Because the key question about retirement is only available in the Canadian census of 1901, the misclassification-corrected-probit approach is arguably the better way to extend our analyses before and after 1900.

Without a resolution to the fundamental data problem, economists cannot produce a reliable pre-Social Security labor force participation series nor clarify these investigations of economic activity among older men before public pension provision. Such a resolution would permit a clearer understanding of their labor force behavior, including assessment of factors that led men to withdraw from the labor force, and would facilitate a more accurate measurement of the impact of Social Security itself. What these desiderata require is a reliable LFPR series before 1940, an adjustment of gainful employment data to make them equivalent to labor force data. Earlier approaches relied on arbitrary assumptions about unemployment levels, or decisions as to what constituted a legitimate gainful occupation. More critical to the solution is attention to the core inquiry in labor force questions, i.e., whether the respondent is actively seeking work.

This paper makes significant strides towards that end. By extending to older men in the US a unique feature of the 1901 Canadian census in which respondents were asked their retirement status as well as their prior occupations, we find that misreporting of employment status substantially inflated historical rates of labor-force participation, particularly for men 65 and older. We further show that using a formal econometric correction for such misclassification produces adjusted rates of labor-force participation that are strikingly similar to those based on the Canadian census, suggesting that both methods produce something close to the true participation rates. We extend our econometric correction to the 1880, 1910, 1920 and 1930 US censuses. Our estimates show that labor-force participation rates for older men based on reported occupations overstate both the true rates and the decline in those rates between the 1900 and 1910 censuses.

The LFPR series produced by this econometric technique avoids potentially arbitrary classifications of occupations and non-occupations. It is lower than older, unadjusted series, but the downward trend is not eliminated. Withdrawal from the labor force, perhaps retirement in the modern sense, was increasing before the establishment and spread of Social Security and other government-provided old age relief programs. The mutual fidelity of our approaches to correction for misclassification of employment status show that both are promising avenues for work on topics where the question of the historical labor-force participation rate is an important one. Future work should systematically explore the differences between, and different implications of, our estimates and those produced elsewhere.

References

Carter, Susan and Richard Sutch, "Fixing the Facts," Historical Methods, (Winter 1996).

Caudill, Steven B, and Franklin G Mixon. 2005. "Analysing Misleading Discrete Responses: A Logit Model Based on Misclassified Data." *Oxford Bulletin of Economics and Statistics* 67 (1): 0305–9049.

Costa, Dora. The Evolution of Retirement: An American Economic History, 1880-1990. Chicago: University of Chicago Press (1998).

Costa, Dora. "Agricultural Decline and the Secular Rise in Male Retirement Rates," Explorations in Economic History, 32 (1995), 540-552.

Darroch, Gordon and Lee Soltow. Property and Inequality in Victorian Ontario: Structural Patterns and Cultural Communities in the 1871 Census. Toronto: University of Toronto Press, 1994.

Dillon, Lisa. "Boundaries of Age: Exploring the Patterns of Young-Old Age among Men, Canada and the United States, 1870-1901," in Eric W. Sager and Peter Baskerville, eds. Household Counts: Canadian Households and Families in 1901. Toronto: University of Toronto Press, 2007, 302-336.

Dillon, Lisa, Brian Gratton, and Jon Moen. "Retirement at the Turn of the Twentieth Century: a Canadian Perspective". Working paper 2009, submitted to the Canadian Historical Review.

Gratton, B. The Politics of Dependency Estimates: Social Security Board Statistics, 1935-1939. Journal of Gerontology: Social Sciences 52B (1997): S1170-S1214.

Gratton, Brian, "The Poverty of Impoverishment Theory: The Economic Well-Being of the Elderly, 1890-1950," Journal of Economic History, 26 (March 1996), 39-61.

Gratton, Brian and Myron Guttman. “Emptying the nest: Older Men in the United States, 1880 to 2000”. Working paper, Arizona State University and the University of Michigan.

Haber, C. and B. Gratton (1994). Old Age and the Search for Security: An American Social History. Bloomington: Indiana University Press.

Hausman, J A, Jason Abrevaya, and F M Scott-Morton. 1998. “Misclassification of the Dependent Variable in a Discrete-Response Setting.” Journal of Econometrics 87: 239–69.

Higgs, Robert (1980). Competition and Coercion.

Lee, Chulhee, “Long Term Unemployment and Retirement in Early Twentieth Century America,” Journal of Economic History, September (1998b), 844-856.

Lee Chulhee, “The Rise of the Welfare State and the Labor Force Participation of Older Males: Evidence from the Pre-Social Security Era,” American Economic Review, May (1998a), 222-226.

Lee, Chulhee, “The Sectoral Shift and the Labor Force Participation of Older Men in the United States, 1880-1940,” Journal of Economic History, (June 2002), 512-523.

Margo, Robert, “The Labor-Force Participation of Older Americans in 1900: Further Results,” Explorations in Economic History, 30 (1993), 409-423.

Moen, J. "The Labor of Older Men: A Comment," Journal of Economic History, 47 (September 1987), pp. 761-767.

Moen, J., “From Gainful Employment to Labor Force: Definitions and a New Estimate of Work Rates of American Males, 1860-1980,” Historical Methods, 21 (1988), 149-159.

Moen, J., “The Unemployment and Retirement of Older Men: Further Evidence from the 1900 and 1910 Censuses,” Historical Methods, 27 (1994), 40-46.

Ransom, Roger and Richard Sutch, “The Labor of Older Americans: Retirement On and Off the Job, 1870-1937,” The Journal of Economic History, (March 1986), 1-30.

Steven Ruggles, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. 2015.Integrated Public Use Microdata Series: Version 6.0 [Machine-readable database]. Minneapolis: University of Minnesota.

Woollard, Matthew. “The Employment and Retirement of Older Men, 1851-1881: further evidence from the census.” Continuity and Change, Vol. 17, No. 3 (2002): 437-463.

Table 1: Labor Force Participation Rates,

Men 60 and Older, Canada 1901 and US 1900

	Canada				United States	
Age	N	GE	% retired	LFP adjusted	N	GE
all	10515	0.81	0.16	0.65	11,981	0.83
60 - 64	3636	0.90	0.07	0.83	4,493	0.89
65 - 69	2719	0.85	0.11	0.74	3,092	0.84
65+	6879	0.76	0.20	0.56	7,488	0.71
70+	4160	0.71	0.27	0.44	4,396	0.62

Source: Dillon, Gratton, and Moen 2010.

Table 2: Adjusting US Labor Force Participation Rates

Canadian 1901 Logit Variables	Logit Coefficients	US Means
Intercept	4.9497	
Age	-0.1009	71.8553
White	0.7108	0.9087
Farm Household	1.0015	0.4352
Rural, nonfarm Hh	-0.0845	0.2704
Native	-0.0011	0.6819
Head	2.0545	0.7566
Literate	0.1818	0.8312
Married	-0.1479	0.6336
Number of Children	0.1645	1.0905
Log Odds		0.5489
Estimated US LFP		0.7144
Adjusted US LFP		0.6339

Notes: Adjusted LFP is for the population of men aged 65 and up.

Table 3: Adjusting US Labor Force Participation Rates

Canadian 1901 Logit Variables	Logit Coefficients	US Means	US Means, No Cotton	US Means, Canadian Border	US Means, No Blacks
Intercept	4.9497				
Age	-0.1009	71.8553	71.8457	71.9337	71.8186
White	0.7108	0.9087	0.9610	0.9867	0.9944
Farm Household	1.0015	0.4352	0.4041	0.3809	0.4283
Rural, nonfarm Hh	-0.0845	0.2704	0.2744	0.2732	0.2645
Native	-0.0011	0.6819	0.6406	0.5997	0.6521
Head	2.0545	0.7566	0.7469	0.7399	0.6521
Literate	0.1818	0.8312	0.8781	0.8977	0.8947
Married	-0.1479	0.6336	0.6283	0.6341	0.6330
Number of Children	0.1645	1.0905	1.0630	1.0411	1.0812
Log Odds		0.5489	0.5404	0.5115	0.4030
Estimated US LFP		0.7144	0.6958	0.6792	0.7022
Adjusted US LFP		0.6339	0.6319	0.6252	0.5994

Notes: US LFP rates are for populations of men 65 and up.

Table 4: Age-group specific labor force participation rate adjustments

	Age group	Reported	Adjusted	N
All	60+	0.77	0.73	126,681
	60-64	0.89	0.86	46,083
	65-69	0.83	0.77	33,862
	70+	0.62	0.50	46,736
No cotton	60+	0.75	0.72	93,725
	60-64	0.87	0.85	33,517
	65-69	0.80	0.76	25,071
	70+	0.59	0.49	35,137
Canadian border	60+	0.75	0.72	45,560
	60-64	0.88	0.85	16,282
	65-69	0.81	0.77	12,121
	70+	0.60	0.49	17,157
Whites	60+	0.76	0.74	115,132
	60-64	0.88	0.87	41,606
	65-69	0.82	0.78	30,999
	70+	0.60	0.51	42,527

Notes: Adjustments created by inserting age-group-specific US means into the logit model presented in Table 3. Group-specific means of observable characteristics based on the larger 5% extract of the 1900 US census.

Table 5: Probit estimates of labor-force participation status

	Probit	Misclassification-corrected
Age	-0.06	-0.07
	(0.00)	(0.00)
White	-0.58	-0.68
	(0.02)	(0.03)
Farm	0.67	0.85
	(0.01)	(0.02)
Rural, nonfarm	-0.17	-0.19
	(0.01)	(0.01)
Native	0.08	0.09
	(0.01)	(0.01)
Head	1.00	1.29
	(0.01)	(0.02)
Literate	0.14	0.16
	(0.01)	(0.02)
Married	0.05	0.07
	(0.01)	(0.01)
Number of children	0.00	0.00
	(0.00)	(0.00)
Constant	4.19	4.71
	(0.05)	(0.07)
α		0.22
		(0.01)
N	12,681	12,681

Notes: Estimated using IPUMS 5% extract of 1900 Census for men aged 60 and older. Dependent variable is an indicator for labor-force participation. α denotes the probability of reporting labor force participation conditional on nonparticipation. Standard errors in parentheses.

Table 6: US labor-force participation rates among US men aged 60 and older in 1900

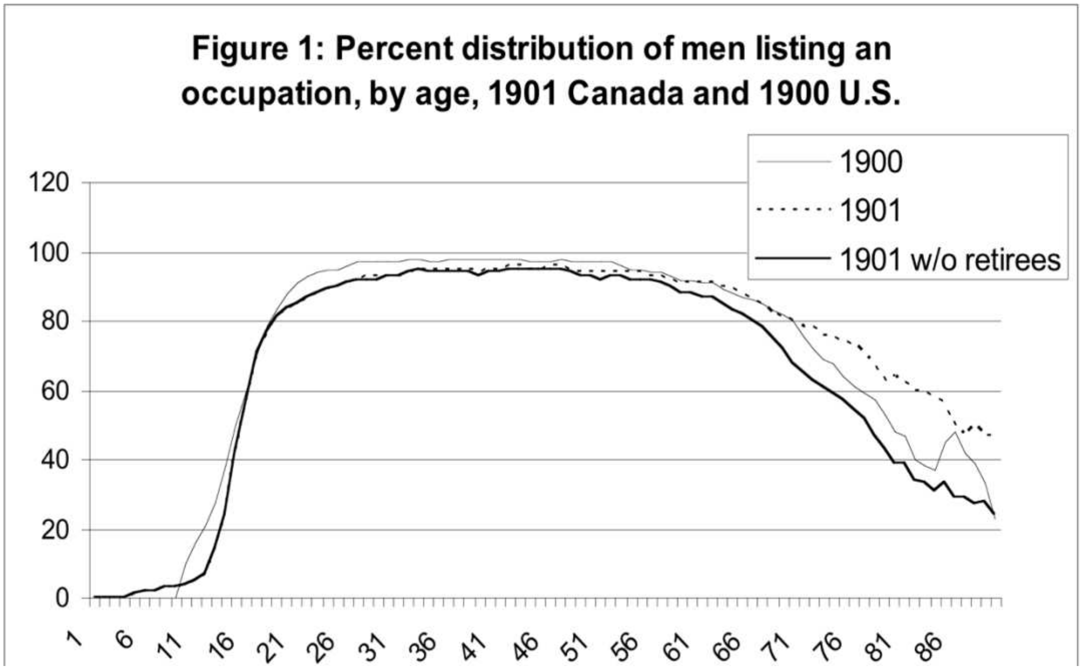
Age group	Reported	Misclassification-corrected prediction	N
All	0.77	0.71	126681
60-64	0.89	0.85	46083
65-69	0.83	0.77	33862
70+	0.62	0.53	46736

Notes: Predicted participation rates are age-group average predicted probabilities of labor force participation, derived from the appropriate probit models.

Table 7: US labor-force participation rates among US white and black men aged 60 and older, 1880-1930

		White			Black		
		Reported	Corrected	N	Reported	Corrected	N
1880	All	0.81	0.69	12,950	0.91	0.89	1,586
	60-64	0.91	0.84	5,115	0.97	0.96	704
	65-69	0.85	0.74	3,364	0.95	0.92	339
	70+	0.66	0.49	4,471	0.82	0.79	543
	α	0.39			0.21		
	SE	(0.02)			(0.16)		
1900	All	0.76	0.69	115,132	0.88	0.88	11,549
	60-64	0.88	0.84	41,606	0.95	0.95	4,477
	65-69	0.82	0.75	30,999	0.93	0.92	2,863
	70+	0.60	0.50	42,527	0.78	0.78	4,209
	α	0.23			0.03		
	SE	(0.01)			(0.07)		
1910	All	0.68	0.63	29,023	0.86	0.79	2,777
	60-64	0.85	0.83	10,817	0.95	0.92	1,130
	65-69	0.72	0.69	7,879	0.91	0.85	690
	70+	0.45	0.39	10,327	0.72	0.60	957
	α	0.11			0.33		
	SE	(0.01)			(0.07)		
1920	All	0.69	0.66	37,765	0.83	0.80	3,173
	60-64	0.86	0.85	14,472	0.93	0.92	1,244
	65-69	0.77	0.72	10,017	0.89	0.86	834
	70+	0.45	0.40	13,276	0.66	0.63	1,095
	α	0.10			0.13		
	SE	(0.01)			(0.06)		
1930	All	0.67	0.65	49,228	0.81	0.78	3,552
	60-64	0.86	0.85	17,799	0.93	0.91	1,429
	65-69	0.75	0.72	13,329	0.87	0.84	931
	70+	0.44	0.41	18,100	0.63	0.57	1,192
	α	0.06			0.15		
	SE	(0.01)			(0.05)		

Notes: Predicted participation rates are age-group average predicted probabilities of labor force participation, derived from the appropriate probit models. "Probability" denotes the probability of reporting labor force participation conditional on nonparticipation. For 1900, estimates are based on a 5% IPUMS Census sample; for remaining years, estimates are based on 1% IPUMS Census samples.



Source: Dillon, Gratton, Moen 2010.