The Effects of Health Shocks and Economic Factors on Retirement Plans Using Quantile Regression Methods for Identifying Unobserved Heterogeneity

By Debra Sabatini Dwyer

Updated, January, 2003

Abstract: Research to date on deviations from plans to retirement suggests that health and economic factors are important predictors of changes to plans (Benitez-Silva and Dwyer 2003, Dwyer, 2002, Dwyer and Hu, 1999). A story one could tell from these findings is that some people plan better than others. Methods used to date have been restricted to conditional mean functions so that we only observe what happens on average. This means we restrict the effects of health and economic factors to be constant for all. The purpose of this work is to address the potential for unexplained factors driving deviations from plans, that might be correlated with an ability to process information to plan well for retirement. Specifically I will use quantile regression analysis on all five waves of the Health and Retirement Study to examine how important factors influence retirement plans conditional on propensity to work.

This proposal is consistent with Social Security priority research areas 2 and 7. I am requesting $25000 for this project.

Introduction

Research to date on deviations from plans to retirement suggests that health and economic factors are important predictors of changes to plans. We find that health plays a bigger role in determining who would like to retire, and that health shocks trigger retirement earlier regardless of whether or not it was planned (Dwyer, 2002, Dwyer and Hu, 1999). Another finding from these studies is that economic factors play a bigger role in determining the retirement outcome then in predicting the plans. One story one could tell from these findings is that some people plan better than others. In other words, health alters tastes for work so that retirement is preferred regardless of economic constraints. We hypothesize that some plan using full information and some do not. Methods used to date have been restricted to conditional mean functions so that we only observe what happens on average. This means we restrict the effects of health and economic factors to be constant for all. As a consequence the reduced form estimators are inefficient and we are unable to identify these "types" of planners without further analysis. Mean-based estimators place a lot of weight on outliers. While this body of research contributes to our understanding of the importance of health and economic status in retirement planning and outcomes, the findings are limited. The purpose of the present study is to obtain estimates not only based on conditional means but over the complete distribution of plan deviations using semi-parametric techniques. In other words, we allow health and economic shocks to have differential effects on various groups of workers by type (related to size of deviation of plan and direction). Dwyer and Benitez-Silva (2002) use panel data techniques to examine unobserved heterogeneity in changes to retirement plan models. They use traditional random effects models with selection that are conditional mean based approaches.

In addition to identifying differential effects of health and economic incentives on deviations to plans between the planning horizon and the outcome stage, we plan to examine the role of these incentives over time using a hazard model approach. By doing so we use full information between the planning horizon and the final stage. In this framework we not only account for the role of important factors in explaining changes to plans at one time, but how the plans evolve dynamically over time. By combining semi-parametric techniques in the hazard framework, we allow for heterogeneity in the model of work duration.

The work speaks to priority research area (1) - the impact of program rules on individuals' work and retirement decisions. We model work and retirement both in the static and dynamic framework. The quantile regressions enable us to examine effects differently for those more tied to the Social Security system (retirement at eligible ages) and those not. This could explain why they are more tied and can inform policy better in terms of equity issues. It also speaks to priority research area (6) in that the quantile regression will enable us to identify unobserved heterogeneity that is likely correlated with income and education. We would like to identify the size of the group classified as in poor health but unable to retire early for financial reasons. This group is likely to experience a big positive deviation from their expected retirement age and more likely to be tied to the public system. Changes to eligibility ages would have bigger adverse consequences for these groups relative to others. In sum, we are still interested in the timing of retirement, accounting for the role of plans in that outcome. We are interested in who plans well and who does not, and what the implications are for policy.

Background and Contributions

Dwyer (2002) uses waves 1 and 4 of the Health and Retirement Study (HRS) to examine the role of health shocks and other incentives in explaining deviations from plans to retirement. Using the static life cycle model of retirement behavior she models plans to retire and then actual retirement as a sequential process. She finds that plans significantly predict outcomes and that individual and spouse health shocks lead to retirement earlier than planned for all. Health explains more of the variation in retirement outcomes for retirements that were not planned. Poor health is less likely to be a shock for those who planned to retire based on existing problems at baseline. In that research workers are categorized based on whether or not they planned to fully retire by wave 4. The analysis separately identifies effects on retirement outcomes conditional on those plans, allowing for the two to be correlated. Economic factors and the availability of health insurance all play important roles. She also allows for a gradual departure from the labor force using an ordered probit of retirement that allows for partial retirement. From this analysis we get a better understanding of different types of plans and outcomes. Finally she examines deviations in a dynamic framework by using a fixed effects (first differences) approach that removes within person unobserved heterogeneity. From this research we learn that both functional status and illness contribute to planning and retirement behavior and how these phenomena interact on average. This analysis will replicate this work allowing for differential effects for different types of planners. In other words, the quantile regression approach will allow us to measure effects differentially based on the propensity to plan well and to retire early.

Retirement analysis restricted to conditional mean estimation may be limited since it is influenced by peaks of retirement at ages 62 and 65. In other words, the distribution may not be normally distributed around some mean. A conditional median approach would probably make more sense. The semi-parametric approach advocated here would allow for not only conditional median estimation, but a complete labor supply estimation function across years of work. The approach was first introduced by Huber (1967) and Koenker and Bassett (1978), where you can find a complete description of the approach. Since then many interesting applications have utilized the approach successfully. Some examples include a more complete assessment of the Pennsylvania Reemployment Bonus Experiments (Koenker, Bilias; 2001); and a more detailed analysis of the gender wage gap by Garcia, Hernandez Lopez-Nicolas' (2001). Of relevance for this research is a paper by Ribeiro (2001) who uses quantile regression analysis with selection to get more complete labor supply elasticities by propensity to work. He finds that income and wage elasticities at the standard work week are zero and negative for those working longer. Ribeiro control for selection based on whether or not you are in the labor force. Conditional mean estimators would have come up with elasticities based on averages across these work groups and therefore tend to be less robust. We will apply this approach analyzing retirement ages controlling for selection into planning to retire and not. What we can learn are the magnitudes of the health and economic factors in explaining retirement outcomes for those who retire at the early age, the normal age, and other ages. We also examine the role of baseline plans differentially by these various groups. The quantiles allow us to separate workers into types based on their retirement outcomes.

One of the limitations of that research is an inability to use information between the planning horizon and the outcome stage. Economic factors at the time of retirement are endogenous and therefore only included at baseline. Baseline expected retirement benefits may be reasonable indicators of future benefits, but wealth and other income could experience shocks that we cannot include. Forni (2002) uses self-reports of wealth and economic shocks but these are noisy and do not perform well in the analysis. The first part of this project focuses on what explains changes to plans, an interesting question. We also propose to answer a slightly different question that enables us to focus in on the effects of evolving economic factors by examining the hazard of working in each period up until wave 5. Again, we plan to allow for different types of planners in this analysis as well. Koenker and Bilias (2001) use quantile regression techniques in a duration model of a Pennsylvania Reemployment Bonus Experiment. They were able to focus attention on particular regions of the conditional duration distribution which we would like to be able to do with retirement given the incentives to retire at various ages under the Social Security system.

In sum, the contributions of this proposed work are (1) to re- examine retirement behavior and planning allowing for differential impacts of health, economic, and socio-economic status for different types of planners and (2) examine the hazard of work departure for potential retirees by health and economic status, again allowing for unobserved heterogeneity in retirement behavior.

Conceptual Retirement Model and Empirical Specifications

In the first part of the analysis, we use the static life cycle model where the retirement decision is based on health status and lifetime earnings. The conceptual model is the same one used in Dwyer (2002) and Dwyer and Hu (1999). Agents choose when to retire or the quantity of retirement leisure, R, and desired consumption, C, which is a function of work. So that they get positive utility from R and C and the utility function is assume to be strictly quasi concave.

We derive an econometric specification where we model R as a function of health, economic, and socio-economic characteristics of the individual and household.

We estimate using Quantile Regression because there are likely asymmetries in the retirement age response to economic incentives, particularly given kinks in the budget line imposed by the Social Security Old Age and Survivors Insurance defined benefit plan. Semi-parametric estimation will allow us to examine effects based on the median age of retirement as well as other important age breakdowns to get at different types of potential retirees. One of the primary variables of interest is the expected retirement age. We want to know how well this variable does in explaining the retirement outcome for different groups of respondents in the data. Because this variable is clearly endogenous, we implement instrumental variables regression using only baseline characteristics in the first stage explaining the expected age and new information in the second stage. We then interpret the coefficients on health, economic, and socioeconomic status as conditional on when they planned to retire. In other words, we can examine shocks because the plans presumably pick up effects of unchanged factors. Like Dwyer (2002) we can examine the role of plans and how important they were in determining outcomes, and also how important changes to the relevant factors are. We can now do this for different types of workers.

In the OLS framework we estimate:

$$R = \alpha_0 + \alpha_1 H + \alpha_2 W + \alpha_3 S + e_1$$

Where R = the retirement age
H = vector of health status indicators of respondent and spouse
W = vector of economic variables
S = vector of socioeconomic factors
e_1 = random error

and α_I for I=1,2,3 is the least squares estimator that fits a line by minimizing the sum of square residuals and estimates the mean of R conditional on the values of H, W, and S.

In the quantile regression framework we estimate:

$Q_\theta(R) = \beta_0 + +\beta_1 H + \beta_2 W + \beta_3 S + e$

Where $Q_\theta(R)$ represents a regression quantile of retirement age. For the median regression $\theta = 0.5$ and we can vary it to specify different quantiles. The median regression minimizes absolute residuals rather than the sum of the squares of residuals and the estimates the median of R conditional on values of H,W, and S. If we specify $\theta = 0.25$ then we estimate the 25th percentile or the first quartile.[1] We can specify any value for θ between 0 and 1 so that we can target analysis for specific retirement ages. Standard errors are corrected following Koenker and Bassett (1982) and Rogers (1993).

In the second part of the analysis we use the standard hazard model framework where we follow workers into retirement and examine the role of health, economic, and socioeconomic status on the duration of work. Again, we do this in the semi-parametric framework following Koenker and Bilias (2001) to get a more complete picture of retirement and work.

Data Used in the Analysis

For the first part of the analysis we use waves 1 and 5 of the HRS. There are roughly 4,000 male and female respondents who provide all information necessary for the analysis in both waves. We use the age they plan to fully retire in wave 1 and the age they reported retirement in wave 5. For a discussion of health, economic, and socio-economic variables see Dwyer (2002). We plan to take advantage of the confidential earnings and benefits data available through the HRS. This prospect may be more easily facilitated through a proposed

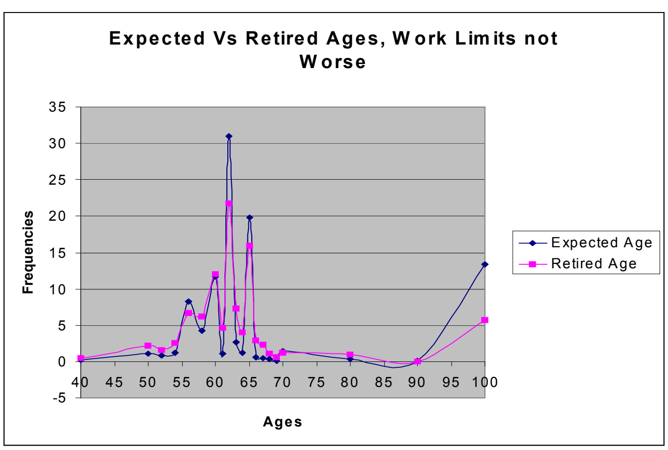
Research Data Center housing the data in Manhattan.[2] For the hazard framework we will use full retirement, health, economic, and socio-economic data from waves 1 through 5.

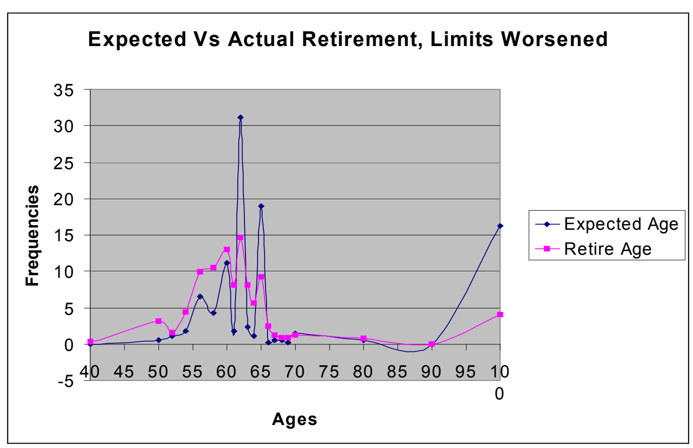
The mean age of retirement in wave 5 is 61 and the median is 62. The first quartile age is 59 and the 3rd is 64. The mean expected age of retirement in wave 1 is 65.5 with a median of 62. The first quartile age is 60 and the 3rd quartile age is 65. People who planned to work by wave 4 were more likely to do so then those who planned to retire. This can be seen in Table 1 below taken from the Dwyer (2002) work. The figures that follow show that retirement ages peak at 62 and 65 and are skewed to the left of those ages. Quantile regression will allow us to examine the accuracy of plans (conditional on new information) for different groups from this distribution.

Table 1. Frequency of Retirement Status in Wave 4, Given Expectations of Retirement Status reported in Wave 1.

	Expected to Retire by Wave 4 n=2,084[3]	Did not expect to retire by Wave 4, n=2,318[4]
Working	727 (34.5%)	1,792 (77.3%)
Partially Retired	449 (21.5%)	237 (10.2%)
Fully Retired	908 (43.6%)	289 (12.5%)

Source: Dwyer, 2002, "Planning for Retirement: The Role of Health Shocks"





[1] So we calculate absolute differences between R and the first quartile value of R (as opposed to the unconditional median which is used in the median regression approach).
[2] SUNY Stony Brook is involved in a proposal to start such a center at CUNY Baruch in Manhattan. As a member we will have access to the data through that forum. The data is not likely to be available until later this year.
[3] 380 missing values for work status in wave 4.

Simply regressing retirement age in wave 5 against the expected age in wave 1 yields a statistically significant coefficient of 0.062, or an increase in the expected retirement age of 1 results in an increase in the actual retirement age of 0.06. Using quantile regression analysis we see that the results are larger in magnitude and stronger for the first and third quartiles then for the mean or median. In other words, if you are in the retirement age ranges of 59 or 64 then the expected age does a better job predicting the actual retirement age then if you are in the mean/median range. This is extremely preliminary but motivates further analysis.

Concluding Comments

The project proposed contributes to the literature on retirement expectation behavior and how it contributes to actual retirement outcomes. It addresses the role of changes to relevant incentives while allowing for unobserved heterogeneity in retirement planning and outcomes. The data are created for the most part. What remains to be done over the next year are refinement and implement of the methodologies. I am requesting $25,000 for this one year project.

[4] 418 missing values for work status in wave 4.

References

Bernheim, B.D. (1987). How DO the Elderly Form Expectations? An Analysis of Responses to New Information, in D.Wise (Ed.), Issues in the Economics of Aging, University of Chicago Press, 259-285.

Blau, D.M., D.B. Gilleskie, and C. Slusher (1997). The Effects of Health on Employment Transitions of of Older Men. Presented at the 1997 annual meeting of the Population Association of America, Washington, DC.

Bound, J., M. Schoenbaum, T.R. Stinebricker, T. Waidmann (1998). The Dynamic Effects of Health on the Labor Force Transitions of Older Workers, NBER Working Paper #6777.

Buchinsky, M. (2001), "Quantile Regression with Sample Selection: Estimating Women's Return to Education in the US", Empirical Economics, V26, #1, pp. 87-113.

Dwyer, D., "Planning for Retirement: The Role of Health Shocks", Unpublished manuscript. SUNY Stony Brook.

Dwyer, D. and H. Benitez-Silva, "Dynamic Retirement Expectations Using the HRS", manuscript. SUNY Stony Brook.

Dwyer, D.S., and O.S. Mitchell (1999). Health Shocks as Determinants of Retirement: Are Self-Rated Measures Endogenous?, Journal of Health Economics, 18:173-193.

Dwyer, D.S. and J. Hu. (1999). The Relationship Between Retirement Expectations and Realizations: The Role of Health Shocks, Presented at the 1998 Meeting of the Pension Research Council, Philadelphia, PA. Forthcoming in conference volume.

Fields, G. O.S. Mitchell (1984). The Economics of Retirement Behavior. Journal of Labor Economics, 2:84-105.

Forni, Lorenzo, (2002) "Expectations and Outcomes", Unpublished Manuscript.

Garcia, J, PJ Hernandez, A Lopez-Nicolas (2001), "How Wide is the Gap: An Investigation of Gender Wage Differences Using Quantile Regression", Empirical Economics, V. 26 #1, pp. 149-167.

Gatti, R. (1997). Savings and Health Shocks. Unpublished Manuscript. Washington, DC.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. Journal of Political Economy. 80(2):223-255.

Irelan, L.M. (1976). Almost 65: Baseline Data from the Retirement History Study. Research Report No. 49, Office of Research and Statistics, US Social Security Administration, Washington, DC. GAO.

Juster, F.T. and R.M. Suzman (1995). An Overview of the Health and Retirement Survey. Journal of Human Resources, 30: Supplement: S7-S56.

Koenker, R and G Bassett (1978), "Regression Quantiles", Econometrica, 46:33-50.

Koenker, R and Y Bilias, (2001), "Quantile Regression for Duration Data: A Reappraisal of the Pennsylvania Reemployment Bonus Experiments", Empirical Economics, V26 #1, pp. 199-220.

Madalla, G.S. (1983). Limited Dependent and Qualitative Variables in Econometrics. Cambridge University Press,

Ribeiro, EP, (2001), "Asymmetric Labor Supply", Empirical Economics, V26 #1, pp. 183-197.

Smith, J.P. (1997), The Changing Economic Circumstances of the Elderly: Income, Wealth, and Social Security. Syracuse University Policy Brief, No. 8.