Quantile regression with sample selection: Estimating women’s return to education in the U.S.*

Moshe Buchinsky
Department of Economics, Brown University, National Bureau of Economic Research and CREST-INSEE, Providence, RI 02912, USA

Abstract. This study uses quantile regression techniques to analyze changes in the returns to education for women. The data used is the March Current Population Survey for the years 1968, 1973, 1979, 1986 and 1990. The first step in estimating the single (linear) index selection equation uses Ichimura’s (1993) semiparametric procedure. To correct for an unknown form of a sample selection bias in the quantile regression, the second step incorporates a nonparametric method, using an idea similar to one developed by Heckman (1980) and Newey (1991) for mean regression, and Buchinsky (1998) for quantile regression.

The results show that: (a) the returns to education increased enormously for the younger cohorts, but very little for the older cohorts; (b) in general the returns are higher at the lower quantiles in the beginning of the sample period and higher at the higher quantiles by the end of the sample period; (c) there is a significant sample selection bias for all age groups at almost all quantiles; (d) toward the end of the sample period there is a significant convergence of the returns at the various quantiles, especially for the younger cohorts and age groups; and (e) the semiparametric estimates of the selection equation are considerably different from those obtained for a parametric probit model.

* This paper was presented at the conference on “Economic Applications of Quantile Regression” held in Konstanz, Germany, June 2000. I wish to thank Don Andrews, Josh Angrist, Claudia Goldin, Jenny Hunt, Larry Katz, Jacob Mincer, Ariel Pakes, Jean-Marc Robin, John Rust, Steve Stern, Joe Tracy, and Joel Waldfogel, as well as seminar participants in Columbia University, Yale University, University of Montreal, University of Virginia, and the Hebrew University of Jerusalem for comments and discussions. I have benefited greatly from insightful comments on earlier drafts by Gary Chamberlain, Jim Powell, James Heckman, Roger Koenker, Bernd Fitzenberger, and an anonymous referee. I am, nonetheless, fully responsible for any remaining errors in this paper.
Key words: Quantile Regression, Nonparametric Selection Correction, Return to Education.

JEL classification: C13, C14, J16, J24, J31.

1. Introduction

The U.S. economy and its position in the world witnessed rapid changes during the 1970’s, and more so during the 1980’s. One important aspect of these changes is the effect on the wage structure of U.S. workers. Much recent research in labor economics documents the basic facts and offers possible explanations for the observed phenomena. But economists, and particularly labor economists, still largely disagree as to the relative importance of the factors contributing to the changes in the U.S. wage structure. Although most studies in this area have concentrated on the male labor force, researchers have not ignored the female labor force. Nevertheless, the focus of research in this area is somewhat different. There is a large body of literature on women’s participation in the labor force, labor supply of women, wage gaps between women and men, and technical issues of sample selection.

The growing importance of the female labor force goes far beyond higher participation. It has long been documented that women have also taken positions requiring more responsibilities and skills. Traditional family and household duties still place more constraints on women than on men, but more married women – and more importantly, married women with young children – pursue independent careers.

This study examines changes in the female wage structure by utilizing quantile regression techniques to analyze changes in the return to education for women using the March Current Population Survey (March CPS) for the years 1968, 1973, 1979, 1986 and 1990. For this purpose, a few new methodological features are introduced. Sample selection bias is corrected in a nonparametric fashion using a two-step procedure first suggested by Heckman (1980), and later developed by Newey (1991), for mean regression.

I examine changes in the return to education for nine different age groups between 20 and 64 years, as well as for each cohort that can be followed throughout the sample period. Specifically, the mean return to education, along with the returns at five quantiles, are estimated for each age group.

The study sheds light on important issues relating to the labor force, particularly the female labor force. The results portray several important findings. The returns to education increased enormously for the younger cohorts, but very little (if at all) for the older cohorts. In general, the returns were higher at the lower quantiles for the younger age groups at the beginning

---


2 See, for example, Gronau (1974), Heckman (1979, 1980) and Mroz (1987). Exceptional in this respect is the study by Goldin (1990) which gives a detailed investigation of many aspects of the female labor force over the last few decades.