Regional Science and Urban Economics 43 (2013) 985–995

Contents lists available at ScienceDirect

Regional Science and Urban Economics

journal homepage: www.elsevier.com/locate/regec

Expected income and labor market choices of US married couples: A locally weighted regression approach☆

Guo Li [a,*], Thomas A. Mroz [b]

[a] Department of Economics, Virginia Tech., Blacksburg, VA, United States
[b] Department of Economics, Clemson University, Clemson, SC 29634, United States

ARTICLE INFO

Article history:
Received 8 February 2013
Received in revised form 24 September 2013
Accepted 25 September 2013
Available online 10 October 2013

Keywords:
Locally weighted regression
Spatial heterogeneity
Expected Income Hypothesis
Labor markets

ABSTRACT

This paper applies a locally weighted scatterplot smoothing (*loess*) method to estimate the spatially heterogeneous wages of demographic groups of workers across precisely defined US labor markets. We estimate a location choice model using data from the National Longitudinal Survey of Youth (NLSY79) using these estimates of labor market specific wages for men and women as determinants of their place of residence. We compare estimates of this model to a model using more aggregated measures of wages and locations from CPS. We show that potential wages based on these more refined definitions of labor markets and demographic groups provide more explanatory power in a simple migration model than do those based upon less detailed definitions of labor markets and demographic groups.

Published by Elsevier B.V.

1. Introduction

Interregional migration is an important economic phenomenon. A number of recent studies have found that it significantly affects spatial equilibrium (e.g., Blanchard et al., 1992; Partridge et al., 2012), local labor market flexibility (e.g., Obstfeld et al., 1998), and industry restructuring (e.g., Dennis and İşcan, 2007). Further, the migration of the skilled workers and the subsequent changes in human capital flows in the local production process has become a major source of regional demographic change and economic growth (e.g., Crescenzi et al., 2007; Faggian and McCann, 2006, 2009; Matano and Naticchioni, 2012). Studies of migration behavior and the subsequent effects on local markets date back to the 1970s with the pioneering Harris–Todaro Model (Harris and Todaro, 1970), where the authors present a two-sector model to show that the unemployment adjusted expected income from the urban area acts as the equilibrating force of the rural–urban migration.[1] There is extensive literature empirically testing this classic theory referred to as “Expected Income Hypothesis”: In the well known recent study of Kennan and Walker (2011), the main conclusion is that the interstate migration decisions are influenced to a substantial extent by income prospects. Several other studies also emphasized the “matching” of a person's skill set to the reward structure of the local labor market as the major driving force of migration (Bayer and McMillan, 2011; Fields, 1979; Lucas, 1997). In general, researchers reached a common notion that those whose skills are most mismatched with the reward structure offered by their current location are those most likely to leave that location; they tend to relocate to locations which offer higher rewards for their particular skills (Borjas and Trejo, 1992; Zakharenko, 2012).

Studies vary considerably in defining the appropriate choice set when specifying the utility function for a worker making migration decisions. Much migration research focuses on interstate moves only (Borjas et al., 1992; Borjas and Trejo, 1992; Ham et al., 2003; Topel, 1986; Kennan and Walker, 2011). Possible reasons for using states to approximate the labor markets include: 1) Data are usually not available about the SMSA or county in which the surveyed person lives; and 2) State boundaries are stable over time and create an exhaustive partition of the United States. However, several studies, e.g., Borjas and Trejo (1992), and Rhode and Strumpf (2003), acknowledge that SMSAs may better approximate local labor markets. Specifically, SMSAs define urban boundaries that serve as the center for local economic activities, and they are used in a number of regional development studies (e.g., Borjas, 1987; McMillen, 1996).

Studies also differ in terms of how to estimate the “expected income” that an individual is faced with when making a migration decision. Because earnings are only observed for the location actually chosen by the individual,[2] it is crucial to obtain a measure of prospective earnings

☆ NIH/NICHD grant (1R01HD047213) provides partial support for this research. We also thank IPUMS.org (Ruggles, Alexander, Genadek, Goeken, Schroeder and Sobek, 2010) for providing harmonized Census/CPS data for this research.

* Corresponding author. Department of Economics, Virginia Tech, Pamplin Hall (0316), Blacksburg, VA 24061.
E-mail address: lgrace@vt.edu (G. Li).

[1] See Gurgand et al. (2012) for a more detailed summary of more recent migration studies.

[2] See the literature in Self-selection and Tiebort sorting, e.g., Tiebout (1956), Bayer (2000), and Bayer and McMillan (2011).

0166-0462/$ – see front matter. Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.regsciurbeco.2013.09.011

offers from each location in the choice set that matches with the potential migrant's skill set. A popular method is to construct a wage based skill measure by using an earnings function that accounts for some factors in the worker's skill set, such as age, years of education, job tenure, union status, marital status, health status, location specific characteristics[3] and time. The form of the earnings function is usually linear which links log of wage/earnings to the control variables, and the spatial heterogeneity across locations is accounted for by fixed effects (Borjas and Trejo, 1992).

A problem with this conventional approach, however, is the possibility of significant variation in local relationships between the regressors and the wage outcome being masked by the global estimates of a specific parametric model (see discussions in Deller, 2011; Paredes and Iturra, 2012; Rhode and Strumpf, 2003). This leads to wage predictions that have little variability across locations and demographic characteristics. This weakness of the wage profiling models to pick up the spatial effects has a direct impact on the performance of the location choice models — with the local effects losing statistical significance. Studies suggest that the definition of the “local area” should be a true approximation of the local labor market which represents the nature of the earnings function. Together with that, the age profiling model should be able to address the spatial heterogeneity issues that arise from local socioeconomic processes and income dynamics. Specifically, McMillen (2010) points out that the fixed-effects approach is only advantageous when spatial effects are constant within discrete, well defined zones.

In this paper we address many of these issues. First, we use *loess* to estimate a nonparametric regression function (see Cleveland and Devlin, 1988; Meese and Wallace, 1991) for technique details). This wide class of smoothing functions accounts for the spatial heterogeneities across labor markets associated with the demographic characteristics of the workers.[4] This loosens the parametric specifications of the earnings function and better fits the data. Second, the estimation of wage offers in a particular labor market uses those observations that are closest to the geographic center of each labor market. Distance measures obtained from the latitude/longitude of the well-defined geographic centers serve as one of several dimensions in the weighting function.[5]

Our contributions in this study also include: 1) using rich Census/CPS data to ensure a large sample space for *loess*[6]; 2) containing all areas in the continental United States in the choice set for modeling the migration decisions of workers; 3) using a Cross Validation method to obtain the optimal bandwidth for *loess*, which helps to mitigate the “curse of dimensionality” discussed in Cleveland and Devlin (1988). Based on the estimations of expected income from *loess*, we set up a location choice model using age 14 and 1994 residence data from the 1979 National Longitudinal Survey of Youth. We construct “personalized” wage offers for each labor market in the choice set using the *loess* wage model, and add in state aggregated wage measures to examine which type of wage offer measures have higher predictive power. States, or US regions, appear to be too broad as units to serve as labor markets, as the state-level average income measures are always insignificant in the location choice model. We also find that using more detailed demographic groups provides more precise estimates.

The remainder of the paper is organized as following: Section 2 discusses the data and the definitions of the geographic boundaries of the local labor markets; Section 3 discusses the locally weighted regression; Section 4 compares the location choice models; and Section 5 concludes.

2. Data and the geographic boundaries of local labor markets

Using SMSAs (Metropolitan and micropolitan statistical areas)[7] as the approximation for the labor markets has several advantages. First, it enables us to distinguish the urban labor markets from the rural labor markets within a state. Second, each SMSA reasonably defines a local economic “market” where employers and employees demand and supply labor. Third, the limited number of SMSA in the US (about 300) do not form an excessively large choice set, and this ensures enough observations in each labor market for more detailed demographic-location groupings.

2.1. Geographic data source and local labor markets

To obtain a benchmark measure of local labor markets in both the urban and the rural areas, we base our geographic boundary definitions on the 1980 Census using definitions provided by http://www.ipums.org/ (see Ruggles et al., 2010). County group (CNTYGP98) serves as components of metropolitan areas (METAREAs) with the boundaries of county groups identified in “1980 County Groups: Detailed Composition and Boundary files”.[8] These 1980 county groups serve as the basis of the consistent public use micro data areas (CONSPUMA) across time. The mappings for these county groups with the “CONSPUMA component file” are available also in IPUMS.[9] The mapping of the above two files gives us the state, metropolitan area and CONSPUMA code for each of the 3171 counties in the USA. We use METAREA as the relevant labor market for counties that are “ürban” (i.e., those that are in a METAREA). We use the CONSPUMAs within each state to define the local labor market for counties that are non-urban (note that the CONSPUMA code is only available in 1980, 1990, and 2000 in the Census data.) Fig. 1 shows the different geographic definitions for urban and rural labor markets. This definition yields 491 US labor markets, with 267 of them being defined as urban and 224 defined as rural.

2.2. Local labor markets and urban/rural centers

The geographic latitude/longitude information for all counties in the US comes from the Census,[10] and we define the 1980 population weighted average latitude/longitude of the counties that constitute a METAREA or a CONSPUMA as the geographic center of the local labor market. We use these centroids to measure distances between labor markets. To illustrate this, take several counties in South Carolina as an example (see Fig. 2). In this map, Greenville County (point D), Spartanburg County (point E) and Anderson County (point F) form a metropolitan area, METAREA 3160: “Greenville–Spartanburg–Anderson SC”, so we define an urban labor market 3160 that represents this METAREA. Let $Latitude_i$ be the latitude of each of the counties in the METAREA, and the pop_i be the population of that county, $tpop_{METAREA}$ be the total population of the METAREA. We calculate the urban center by:

$$LAT_{mean} = \sum_{i=1}^{N_{county}} \frac{pop_{i,1980}}{tpop_{METAREA,1980}} \times Latitude_i$$

[3] Knapp and Gravest (1989) mentioned that the local public goods and amenities are capitalized into wages and housing prices.

[4] Sex, age, race, educational attainment and hours worked per year are considered to be demographic factors affecting income offers for each labor market. Other exogenous factors include local labor market, urban/rural status of the location, and year of employment.

[5] Using geographic distance as a component of the weighting function is related to the literature in Geographically Weighted Estimation. This spatial version of the conditionally parametric model was first used by McMillen (1996) and developed by Brunsdon et al. (1996) and Partridge et al. (2008).

[6] The estimation of wage/earnings offer from each labor market includes ≥500,000 observations for each year.

[7] Also referred to as METAREA here. See Appendix A for a detailed illustration of Census geographic variables used here.

[8] See: http://usa.ipums.org/usa/volii/ctygrp.shtml. Date accessed: Feb. 2009.

[9] See: http://usa.ipums.org/usa/volii/conspuma_components.xls. Date accessed: Feb. 2009.

[10] See: http://www.census.gov/tiger/tms/gazetteer/county2k.txt.