Journal of Housing Economics 29 (2015) 59–71

Contents lists available at ScienceDirect

Journal of Housing Economics

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

Use of a Gini index to examine housing price heterogeneity: A quantile approach ☆

Jaume García Villar [a], Josep Maria Raya [b,*]

[a] Department of Economics and Business, Universitat Pompeu Fabra, Spain
[b] Escola Superior de Ciències Socials i de la Empresa (Tecnocampus, Universitat Pompeu Fabra), C/Ernest Lluch 32, 08302 Mataro (Barcelona), Spain

ARTICLE INFO

Article history:
Received 1 June 2015
Accepted 8 June 2015
Available online 2 July 2015

JEL classification:
R21
C30

Keywords:
Quantile regressions
Housing prices
Hedonic models
Gini index
Heterogeneity in housing prices

ABSTRACT

This paper contributes to the existing literature that deals with the full distribution of house prices and its decomposition (primarily McMillen, 2008) by conducting a deeper analysis of housing price heterogeneity. Our approach differs from McMillen's insofar as our goal is to explain the variation in housing prices at a point in time rather than over a period of time. The basic statistic used to summarise house price distribution is the Gini index, which compares the actual distribution of the price per square metre (PPSM) with a uniform distribution. We decompose the Gini index into what can be explained for by the explanatory variables (which can also be easily decomposed into the contribution of each explanatory variable) and what remained unexplained. With a data set that includes appraisal values for 9297 dwellings in Barcelona in 1998–2001, the part explained by the standard OLS slopes (up to 60%) suggests a high degree of homogeneity in the linkage between PPSM and the explanatory variables. In any event, when heterogeneity is introduced using a quantile approach, that part of the Gini index explained for by the regressors falls. Finally, the variable that produces the most heterogeneity is area.

1. Introduction

Until recently, the housing economics literature has not dealt with the full distribution of house prices and specifically the decomposition of house prices. This paper contributes to the existing literature (primarily McMillen, 2008) by conducting a deeper analysis of heterogeneity of housing prices., McMillen, 2008) analysed changes in the full distribution of housing prices using data for sales of single-family dwellings in Chicago between 1995 and 2005. He found that the price of dwellings at higher percentiles rose faster than other housing and that the 2005 distribution implied a reduction in house price inequality. The main issue addressed by the paper is the relative contribution of the variation in quantile regression coefficients versus variation in the distribution of the variables explaining house price distribution. This paper complents McMillen (2008) 'is the ction section in order to emphasize the differences between the paper and the McMilnts McMillen (2008), who sets out to decompose changes in the distribution of house prices. Our approach differs from McMillen's insofar as our goal is to use Gini indexes to explain the variation in housing prices at a point in time rather than over a period of time. In this sense, unlike McMillen's, our paper is cross-sectional.

McMillen (2008), using a quantile approach, decomposes changes in the distribution of housing prices over time into items arising from: (1) changes in the distribution of the explanatory variables; (2) changes in the

☆ The authors would like to thank the Ministry of Technology and Science for the funding received through Project ECO2008-06395-C05-01.
* Corresponding author. Fax: +34 931696505.
E-mail address: josep.raya@upf.edu (J.M. Raya).

http://dx.doi.org/10.1016/j.jhe.2015.06.001
1051-1377/

distribution of estimated quantile coefficients of the hedonic price function. Quantile Regression (QR) is frequently used when estimation of the conditional mean cannot capture links between the dependent variable and the explanatory variables throughout the whole distribution of the former. The method is commonly used in other fields of economics, particularly in connection with inequality. QR has also recently been used in the literature on housing economics[1]. The approach used by McMillen is closely related to the literature on changes in earnings inequality. In particular, he follows Machado and Mata, 2005. Machado and Mata (2005), who propose a quantile regression-based decomposition based on the estimation of marginal wage distributions consistent with a conditional distribution estimated by quantile regression. Recently, Nicodemo and Raya (2012) present a relevant multi-city analysis.

In both health economics and labour economics, greater attention is paid to the decomposition issues addresses here. Wagstaff et al. (2002) demonstrate how the linear regression model can be used to decompose indices of inequality to identify the relative contributions made by the explanatory variables. This decomposition treats individuals' responses as homogeneous. Jones and Lopez (2002) demonstrate how this approach can be extended to allow individual heterogeneity through the use of quantile regression, producing an additional source of variation: the difference in coefficients across quantiles.

Our data span only four years (1998–2001) because our main aim is to provide a snapshot of the factors determining the housing price distribution during these years. The basic statistic used to summarise house price distribution is the Gini index (a universally used inequality index), which compares the actual distribution of the **price per square metre** (occasionally referred to here as **PPSM** for the sake of brevity) with a uniform distribution. We ranked housing prices from lowest to highest so we could compare the cumulative share with the 45-degree line distribution. We also studied inequality in the distribution of the PPSM variable and its decomposition into what can be explained for by the explanatory variables (which can also be easily decomposed into the contribution of each explanatory variable) and what remained unexplained. Our data set covered appraisal values for 9297 dwellings in Barcelona between 1998 and 2001. We used this to study changes in the Gini index over this short period of time and its decomposition into the portions due to the explained part and the unexplained part. Finally, we used a quantile model to capture information hidden in the unexplained part (unobserved heterogeneity) in order to estimate the full impact of the explanatory variables on the Gini index at any given point in the full price distribution.

The study is structured as follows: Section 2 demonstrates how the Gini index can be broken down into the contributions made by the various explanatory variables, whether these be individual, homogeneous or heterogeneous ones. We then present the empirical model by introducing the econometric methods that allow this decomposition. Section 3 presents the data and Section 4 the results. In the latter section, we also compare estimates based on OLS and on quantile regressions. Section 5 reports the Gini index findings and the impact of the explanatory variables on house prices using the previous estimates. Section 6 summarises the main conclusions of the study.

2. Methodology

The key methodology used in our study is the decomposition of a measure of inequality. We use the Gini index for price per square metre as a value for measuring housing inequality. The relative impacts of the explanatory variables on this measure are identified by means of a linear regression model.

The Gini index is an inequality measure that is usually associated with a descriptive approach to inequality measurement. The Gini index is named after the Italian statistician Corrado Gini, who invented the measure and published it in his 1912 paper *Variabilità e mutabilità.* It is closely linked to the representation of income inequality through the Lorenz curve[2]. In particular, it measures the ratio of the area between the Lorenz curve and the equidistribution line or 45-degree line (the extreme case where all incomes are equal) to the area of maximum concentration (the area between the Lorenz curve of an income distribution where all incomes are zero except for the last one and the equidistribution line).

This geometrical interpretation based on the Lorenz curve is, however, only one way to calculate the Gini index. Another approach, which proves particularly useful below, is to directly express the Gini index in terms of the covariance between the levels of a particular variables and its cumulative distribution (Lambert, 1993). In terms of our objective, we consider p_i to be the price per square metre of dwelling i and R_i the cumulative proportion of dwellings ordered according to p_i up to dwelling i (relative rank). Thus, the Gini index, G, for price per square metre is:

$$G = \left(\frac{2}{\bar{p}}\right)\mathrm{cov}(p_i, R_i) \tag{1}$$

where $\bar{p}$ is the average price per square metre and Cov is the covariance between the price per square metre p_i and the cumulative proportion of dwellings ordered according to p_i up to dwelling i.

Wagstaff et al. (2002) demonstrate how the linear regression model can be used to decompose indices of inequality so that one can identify the relative contributions made by the explanatory variables. Thus, following Wagstaff et al. (2002) if p_i is constructed from the following linear regression model (where ε_i are the residuals):

$$p_i = \beta_1 + \sum_{k=2}^{K}\beta_k x_{ki} + \varepsilon_i \tag{2}$$

[1] See McMillen and Thorsnes (2006), Coulson and McMillen (2007), Zietz et al. (2008) and Liao and Wang (2012).

[2] The Lorenz curve is a graphical representation of the cumulative distribution function of a probability distribution; it is a graph showing the proportion of the distribution taken up by the bottom y% of the values.