Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Comparative Economics 32 (2004) 348–363
Journal of COMPARATIVE ECONOMICS
www.elsevier.com/locate/jce

Accounting for income inequality in rural China: a regression-based approach

Guanghua Wan

World Institute for Development Economics Research, United Nations University (UNU-WIDER), Katajanokanlaituri 6 B, Fin-00160 Helsinki, Finland

Received 26 February 2003; revised 15 February 2004

Available online 15 April 2004

Wan, Guanghua—Accounting for income inequality in rural China: a regression-based approach

This paper proposes a framework for inequality decomposition in which inequality of the target variable, e.g., income, can be decomposed into components associated with any number of determinants or proxy variables in a regression equation. The proposed framework is general enough to be applied to any inequality measure and it imposes few restrictions on the specification of the regression model. This generality is illustrated by quantifying root sources of regional income inequality in rural China using a combined Box–Cox and Box–Tidwell income-generating function. *Journal of Comparative Economics* **32** (2) (2004) 348–363. World Institute for Development Economics Research, United Nations University (UNU-WIDER), Katajanokanlaituri 6 B, Fin-00160 Helsinki, Finland.

JEL classification: O15; C43; C63

Keywords: Inequality decomposition; Box–Cox model; Income-generating function; China

1. Introduction

A resurgence of interest in income distribution in developed, developing and transition economies is found in the literature. Atkinson (2001) notes the doubling back of inequality after an inverted-U pattern in developed countries. Cornia and Popov (2001) investigate rapidly rising inequality in transitional economies. Datt and Ravallion (1992) and Dollar and Kraay (2002) consider the recent controversy over the effects of growth versus redistribution on poverty reduction in the developing world. Traditional approaches

E-mail address: wan@wider.unu.edu.

0147-5967/$ – see front matter
doi:10.1016/j.jce.2004.02.005

to income distribution are mostly descriptive rather than prescriptive. They involve measuring the extent of inequality and speculating on its determinants. Following the work of Shorrocks (1980, 1982), inequality decomposition by income sources requires an identity to express income as the sum of several components. Conversely, inequality decomposition by population subgroups provides rather limited information on the fundamental determinants of inequality, e.g., differences in human and physical capital, dependency ratios, globalization, and technical change.

Since the early 1970s, economists have used the regression-based approach to inequality decomposition. Unlike its traditional counterparts, this approach allows the contributions of the regression variables to total inequality to be quantified. Although the early work is limited regarding the number and type of variables that can be considered, recent advances have relaxed this restriction. In theory, regression-based inequality decomposition permits the inclusion of any number or type of variables or even proxies, including social, economic, demographic and policy factors. The flexibility of this approach, particularly its ability to accommodate endogeneity of income determination and random errors, makes it attractive to economists, and policy-makers. Oaxaca (1973) and Blinder (1973) are the pioneers of this approach; they focus on the difference in mean income between two groups. Other moments of the income-generating process are not considered. Juhn et al. (1993) extend this approach so that the decomposition depends on the difference in the entire income distribution between two groups rather than on the difference in mean income only. Bourguignon et al. (2001) relax the requirement of a linear income-generating function imposed by Juhn et al. (1993). All these authors focus on explaining differences in income distribution between distinct groups of income recipients; they do not quantify the contributions of specific factors to total inequality. Hence, only a limited number of inequality-related impacts can be identified, although these impacts could be functions of more fundamental determinants. For example, the technique proposed by Bourguignon et al. (2001) can be used to decompose differences in income distribution into only three broad components, namely, price effects, participation effects and population effects.

In a different strand of literature using semiparametric and nonparametric techniques, DiNardo et al. (1996) and Deaton (1997) describe and compare the entire distribution of the target variable in terms of the density function, rather than attempting to decompose a summary measure of inequality. Although they impose few structural assumptions, the findings are less conclusive than economists and policy makers would prefer, as Morduch and Sicular (2002) argue. Fields and Yoo (2000) and Morduch and Sicular (2002) employ conventional techniques to specify and estimate parametric income-generating functions and derive inequality decompositions based on the estimated regression equations. Their conceptual framework allows for any number of fundamental income determinants, but suffers from a number of restrictions. Our paper builds on the work of Fields and Yoo (2000) and Morduch and Sicular (2002); we present a critical evaluation of these papers in Section 2.

Our paper is motivated by three issues. First, any regression-based inequality decomposition inevitably involves a constant term and a residual term. These terms give rise to specific problems, which are neglected or not properly addressed in Morduch and Sicular (2002) and Fields and Yoo (2000). Second, the current state of art in regression-based inequality decomposition imposes severe limitations in terms of the functional forms and