Labour Economics 42 (2016) 76–86

Contents lists available at ScienceDirect

Labour Economics

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

A simple diagnostic to investigate instrument validity and heterogeneous effects when using a single instrument☆

Steven G. Dieterle*, Andy Snell

University of Edinburgh, School of Economics, 30 Buccleuch Place, Edinburgh EH8 9JT, United Kingdom

HIGHLIGHTS

- Many papers using instrumental variables use a first stage linear in the instrument.
- We test the sensitivity to using a quadratic in the instrument in the first stage.
- This simple change gives statistically different results for several papers studied.
- We reconcile these differences using a heterogeneous effects interpretation of IV.
- The heterogeneity uncovered can be used as a further check of instrument validity.

ARTICLE INFO

Article history:
Received 31 August 2015
Received in revised form 23 June 2016
Accepted 16 August 2016
Available online 21 August 2016

JEL classification:
C26

Keywords:
Instrumental variables
Heterogeneous treatment effects
Robustness check

ABSTRACT

Many studies that use instrumental variables are based on a first stage linear in the instrument. Using only linear first stages may miss important information about effect heterogeneity and instrument validity. Analyzing fifteen studies using linear first stages, we find ten with significant nonlinearities. Six of these ten have statistically different second stage estimates. Additional analysis is necessary when results are sensitive to first stage choice. We provide a framework to reconcile these differences by determining those patterns of heterogeneity that are consistent with instrument validity. If these patterns violate economic reasoning, then the validity of the instrument is questioned.

1. Introduction

Economists often employ instrumental variable (IV) techniques when faced with the difficult task of estimating causal effects in non-experimental settings. The first order issue is to find plausibly exogenous instruments. Given that the necessary exogeneity assumption is effectively untestable, in most cases instrument validity is argued on heuristic grounds. On top of validity concerns, interpretation of IV estimates is made more difficult by allowing for unmodeled heterogeneity in responses, a concept made popular in economics due to the influential work of Imbens and Angrist (1994) and Heckman and Vytlacil (1999).

While there are many ways to implement an IV strategy, one of the most common among applied economists is to use Two-Stage Least Squares (2SLS) with the first stage linear in a single instrument.[1] However, using only linear first stages may obscure

☆ We would like to thank seminar participants at CMPO Bristol University, Heriot-Watt University, Michigan State University, the Tinbergen Institute, the University of Bath, University College Dublin, and University of York, as well as two anonymous referees for helpful comments. All errors are the responsibilities of the authors. An earlier version of the paper was circulated with the title "It's Hip to be Square: Using Quadratic First Stages to Investigate Instrument Validity and Heterogeneous Effects.".

* Corresponding author.
E-mail addresses: steven.dieterle@ed.ac.uk (S. Dieterle), A.J.Snell@ed.ac.uk (A. Snell).

[1] This focus on linear first stages is understandable given that the properties of the estimator are well understood relative to nonparametric approaches. For instance, Hansen (2009) notes the "worrisome" issue that many nonparametric approaches are "incomplete" due to ambiguity over bandwidth selection, an issue "critical to implementation". In addition, it is closely connected to the counter factual outcomes framework used in program evaluation with binary treatment and instruments. Furthermore, in traditional treatments of IV first stage choice only impacts efficiency and not consistency, while with heterogeneous effects different first stages estimate arbitrarily different weighted average partial effects. Researchers may also be cautious of the "forbidden regression" problem of using fitted values from a nonlinear, say Probit, first stage directly in the second stage (Angrist and Pischke, 2009). Coupled with concerns over weak instruments with overidentification, these considerations make the linear first stage choice appealing.

http://dx.doi.org/10.1016/j.labeco.2016.08.002
0927-5371/

important information on the nature of heterogeneous effects that can, in turn, augment the heuristic arguments made for instrument validity. We argue that the sensitivity of 2SLS estimates to simple changes in the first stage is an important piece of information that should be routinely reported along with other common diagnostics, like the first stage F-statistic. In this paper, we adapt the heterogeneous effects framework in order to characterize and assess previously undocumented dimensions of heterogeneity that result from using different first stage functions of a single instrument.

To start, we identify cases where the results are sensitive to the first stage functional form by following a basic textbook approach to overidentification testing. In particular, we start by extending the first stage to include a squared term in the instrument.[2] We then test for significance of the quadratic first stage relative to the linear. Finally, we test the sensitivity of the 2SLS estimates to the choice of linear or quadratic first stage using a standard overidentification test — treating the squared instrument as an overidentification restriction. Surprisingly, this simple and nearly costless to implement procedure proves to be empirically relevant when applied to papers relying on first stages linear in a single instrument. Across the fifteen papers we study here, we find evidence of significant nonlinearities in ten papers. Six of these ten studies have cases where the significant quadratic first stage is associated with a statistically significant difference in the 2SLS estimates of interest.

The obvious question — and primary focus of this paper becomes: what should we do when our results are sensitive to the choice of functional form for a single instrument in the first stage?

In a classic treatment of 2SLS with homogeneous effects, different functions of the instrument will affect efficiency, but should identify the same population parameter (Angrist et al., 2000; Heckman et al., 2006; Wooldridge, 2010). Therefore, the sensitivity can be cast as evidence of an invalid instrument.[3] Alternatively, the sensitivity may be evidence of unmodeled heterogeneity with different first stages identifying different weighted averages of underlying responses (Angrist et al., 2000; Heckman et al., 2006). Such heterogeneity may come from a number of sources including nonlinearity in the second stage relationship, as well as more complex forms due to non-separable errors, or individual level functional form differences.

We provide a framework for extracting information about potential heterogeneity from using different first stages. Building on prior work by Angrist et al. (2000), we show that the difference in the estimators (linear and quadratic first-stage) is driven completely by applying different weights to the underlying heterogeneous partial effects at different values of the instrument. Furthermore, we show that the weight ratio at each value of the instrument is easily estimated using only the first stage fitted values without imposing any additional assumptions on the most general heterogeneous effect models. Combined with subsample estimation, the weight ratios allow the researcher to infer the relative pattern of the average partial effects across the distribution of the instrument that would be consistent with a valid instrument.

We argue that the pattern of heterogeneity uncovered by our approach should be checked for a reasonable economic explanation. If it can be matched to a sensible economic story, then we can strengthen our understanding of the question being studied. The results may also justify pursuing more complex estimation approaches, such as nonparametric IV (Newey, 2013) or Local IV (Heckman and Vytlacil, 1999), that tackle effect heterogeneity head on. However, if the pattern does not match a sensible economic story, then the results should be interpreted with caution as it raises concerns over the validity of the instrument.

To illustrate the usefulness of the proposed approach, we compare linear and quadratic first stages for two well-published papers relying on continuous instruments for identification: Becker and Woessmann's 2009 paper on the effects of Protestantism on economic prosperity and Acemoglu et al.'s influential 2001 paper exploring the relationship between institutions and growth. We highlight these two papers as, in each case, we find evidence that adding the square of the instrument to the first stage is important for the final estimates. When exploring the heterogeneous effects explanation for Becker and Woessmann (2009), we find that the implied effects actually change sign (from positive to negative) across the instrument distribution suggesting a very important pattern of heterogeneity. Again this pattern should be matched with a sensible economic story to help bolster the argument for instrument validity.

Since the key papers were chosen to illustrate the important conclusions that may be drawn when non-linear first-stages seem to matter, we also present a survey exercise applying our approach to an objectively chosen set of thirteen papers drawn from *American Economic Association* journals. That we find rejections in over half of the papers underscores the importance of applying this approach generally.

We readily note that while the use of nonlinear transformations of instruments is not, in-and-of-itself, novel, our approach is. This paper is the first to compare estimates from different first stages to show how nonlinearity in the first stage can be exploited to enhance the heuristic arguments for instrument choice by uncovering patterns of heterogeneity with respect to the instrument. Importantly, the patterns of heterogeneity uncovered here typically go unnoticed in empirical work. Our approach also compliments recent work by Lochner and Moretti (2011) and Løken et al. (2012) that considers the importance of nonlinear second stages for typical instrumental variable estimators. The key point of distinction here is our focus on using the nonlinearity in the first stage to test the sensitivity of 2SLS estimates.

The paper proceeds as follows: Section 2 discusses the motivation for considering higher order terms in the homogeneous effects setting; Section 3 applies this approach to the two key examples; Section 4 shows how to characterize the weight ratios in a heterogeneous effects framework and applies this to the Becker and Woessmann (2009) example; Section 5 summarizes the literature survey exercise; and Section 6 concludes.

2. Quadratic overidentification test

To motivate our approach, we begin with a simple text-book treatment of instrumental variables. Later, we will consider the implications under a more general heterogeneous effects setting. Following Wooldridge (2010), start with a linear model for y in terms of x in the population:

$$y = \mathbf{x}\beta + u \tag{2.1}$$

where $\mathbf{x} = (1, x_2, \ldots, x_K)$ is a vector of covariates.

Further denote our instrument vector by $\mathbf{z} = (1, x_2, \ldots, x_{K-1}, z)$, where we assume one endogenous regressor (x_K) and a single

[2] While the arguments we make will also hold for higher order polynomials (and other functional forms), we find that the quadratic first stage is sufficient to uncover evidence of nonlinearity in most cases even when higher order terms would improve the fit. Furthermore, by choosing the quadratic first stage we avoid generating weak instrument problems by adding only one overidentification restriction and we have a simple test that can be uniformly applied across cases to avoid data mining.

[3] This interpretation can be extended to more general cases where heterogeneous effects are independent of the instrument (Heckman et al., 2006).