Journal of Econometrics 193 (2016) 162–182

Contents lists available at ScienceDirect

Journal of Econometrics

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

Informational content of special regressors in heteroskedastic binary response models[☆]

Songnian Chen[a], Shakeeb Khan[b,*], Xun Tang[c]

[a] Department of Economics, Hong Kong University of Science and Technology, Hong Kong
[b] Department of Economics, Duke University, United States
[c] Department of Economics, Rice University, United States

A R T I C L E I N F O

Article history:
Received 25 April 2013
Received in revised form
3 December 2015
Accepted 28 December 2015
Available online 27 February 2016

JEL classification:
C14
C21
C25

Keywords:
Binary response
Heteroskedasticity
Identification
Information
Median independence
Conditional symmetry

A B S T R A C T

We quantify the informational content of special regressors in heteroskedastic binary response models with median-independent or conditionally symmetric errors. Based on Lewbel (1998), a special regressor is additively separable in the latent payoff and conditionally independent from the error term. We find that with median-independent errors a special regressor does not increase the identifying power by a criterion in Manski (1988) or lead to positive Fisher information for the coefficients, even though it does help recover the average structural function. With conditionally symmetric errors, a special regressor improves the identifying power, and the information for coefficients is positive under mild conditions. We propose two estimators for binary response models with conditionally symmetric errors and special regressors.

1. Introduction

In this paper we explore the informational content of a special regressor in binary choice models. A special regressor is one that is additively separable from all other components in the latent payoff and that satisfies an exclusion restriction (i.e., it is independent from the error term conditional on all other regressors). In this paper, our definition of a special regressor per se does not require it to satisfy any "large support" requirement. We examine how a special regressor contributes to the identification and the Fisher information of coefficients in semiparametric binary response models with heteroskedastic errors. We focus on the role of special regressors in two models where errors are median-independent or conditionally symmetric, respectively.

Special regressors arise in various social-economic contexts. Lewbel (2000) used a special regressor to recover coefficients in semiparametric binary response models where heteroskedastic errors are mean-independent from regressors. He showed coefficients for all regressors and the error distribution are identified up to scale, provided that the support of special regressor is large enough. Lewbel (2000) then proposed a two-step inverse-density-weighted estimator. Since then, arguments based on special regressors have been used to identify structural micro-econometric models in a variety of contexts. These include multinomial-choice demand models with heterogeneous consumers (e.g. Berry and Haile, 2010); static games of incomplete information with player-specific regressors excluded from interaction effects (e.g. Lewbel and Tang, 2015); and matching games with unobserved heterogeneity (e.g. Fox and Yang, 2012).

Using a special regressor to identify coefficients in binary response models with heteroskedastic errors typically requires additional conditions on the support of the special regressor. For

☆ We are grateful to Brendan Kline, Arthur Lewbel, Jim Powell, Frank Schorfheide and Haiqing Xu for useful feedback. We also thank conference and seminar participants at Cambridge, CEMFI, China ES Meetings, Cornell, CREST, Northwestern, Pompeu Fabra, SETA (Taiwan), Texas Econometrics Camp, Tilburg, UCL, UT Austin and Yale for comments. We thank Bingzhi Zhao, Yichong Zhang and Michelle Tyler for capable research assistance. The usual disclaimer applies.

* Corresponding author.
E-mail addresses: snchen@hst.hk (S. Chen), shakeebk@duke.edu (S. Khan), xun.tang@rice.edu (X. Tang).

http://dx.doi.org/10.1016/j.jeconom.2015.12.018
0304-4076/

instance, in the case with mean-independent errors, identification of linear coefficients requires the support of special regressors to be at least as large as that of errors. Khan and Tamer (2010) argued that point identification of coefficients under mean-independent errors is lost whenever the support of special regressor is bounded.[1] They also showed that when the support of a special regressor is unbounded, the Fisher information for coefficients is zero when the second moment of regressors is finite.

The econometrics literature on semiparametric binary response models has largely been silent about how to use special regressors in combination with alternative stochastic restrictions on errors that require less stringent conditions on the support of special regressors. Magnac and Maurin (2007) introduced a new restriction on the tail behavior of the latent utility distribution outside the support of special regressors. They established the identification of coefficients under such restrictions. The tail condition they used is not directly linked to more conventional stochastic restrictions on heteroskedastic errors, such as median independence or conditional symmetry. (We show in Appendix B that the tail condition in Magnac and Maurin (2007) and the conditional symmetry considered in our paper are non-nested.) We show the information for coefficients is positive in our model under conditional symmetry with a special regressor.

We contribute to the literature on binary choice models with several findings. First, we quantify the change in the identifying power of the model due to the presence of special regressors under median-independent or conditionally symmetric errors. This is done following the approach used in Manski (1988), which involves comparing the set of states where the conditional choice probabilities can be used for distinguishing true coefficients from other elements in the parameter space. For the model with median-independent errors, we find that further restricting one of the regressors to be special does not improve the identifying power for coefficients. For the model with conditionally symmetric errors, we find that using a special regressor does add to the identifying power for coefficients in the sense that it leads to an additional set of (paired states) that can be used for recovering the true coefficients. This is a surprising insight, because Manski (1988) showed that, in the absence of a special regressor, the stronger restriction of conditional symmetry adds no identifying power relative to the weaker restriction of median independence.

Second, we show how the presence of a special regressor contributes to the information for coefficients in these two semiparametric binary response models with heteroskedastic errors. For models with median-independent errors, we find the Fisher information for coefficients is zero even when one of the regressors is special. In comparison, for models with conditionally symmetric errors, we find the presence of a special regressor does yield positive information for coefficients. We also propose two consistent estimators for linear coefficients when errors are conditionally symmetric. These two results seem to suggest that there exists a link between the two distinct ways of quantifying informational content in such a semiparametric model: the set of states that help identify the true coefficients in Manski (1988), and the Fisher information for coefficients in semiparametric binary response models in Chamberlain (1986).

Our third set of results (Section 3.3) provide a more positive perspective on the role of special regressors in structural analyses. We argue that, even though a special regressor does not add to the identifying power or information for coefficients when heteroskedastic errors are median-independent, it is instrumental for recovering the average structural function, as long as the support of the special regressor is large enough.

This paper contributes to a broad econometrics literature on the identification, estimation and information of semiparametric limited response models with heteroskedastic errors. A partial list of other papers that discussed related topics include Chamberlain (1986), Chen and Khan (2003), Cosslett (1987), Horowitz (1992), Blevins and Khan (2013), Khan (2013), Magnac and Maurin (2007), Manski (1988) and Zheng (1995) (which studied semiparametric binary response models with various specifications of heteroskedastic errors); as well as Andrews (1994), Newey and McFadden (1994), Powell (1994) and Ichimura and Lee (2010) (which discussed asymptotic properties of semiparametric M-estimators).

2. Preliminaries

Consider a binary response model:

$$Y = 1\{X\beta - V \geq \epsilon\} \tag{1}$$

where $X \in \mathbb{R}^k$, $V \in \mathbb{R}$ and $\epsilon \in \mathbb{R}$, and the first coordinate in X is a constant. We use upper-case letters for random variables and lower-case letters for their realized values. Let F_R, f_R and Ω_R denote the distribution, the density and the support of a random vector R respectively; let $F_{R_1|R_2}, f_{R_1|R_2}$ and $\Omega_{R_1|R_2}$ denote the conditional distribution, density and support in the data-generating process (DGP); let $F_{R_1|r_2}$ be shorthand for $F_{R_1|R_2=r_2}$ and likewise for $f_{R_1|r_2}$ and $\Omega_{R_1|r_2}$. Assume the marginal effect of V is known to be negative, and set it to -1 as a scale normalization. We maintain the following condition throughout the paper.

Assumption 2.1 (*Special Regressor*)**.** V is independent from ϵ given any $x \in \Omega_X$.

For the rest of the paper, we sometimes refer to this condition as an "exclusion restriction", and use the terms "special regressors" and "excluded regressors" interchangeably. Let Θ be the parameter space for $F_{\epsilon|X}$ (i.e. Θ is a collection of all conditional distributions of errors that satisfy the model restrictions imposed on $F_{\epsilon|X}$). Let $Z \equiv (X, V)$ denote the vector of regressors reported in the data. The distribution $F_{V|X}$ and the conditional choice probabilities $\Pr(Y = 1 \mid Z)$ are both directly identifiable from the data and considered known in the discussion about identification. Let $p(z)$ denote $\Pr(Y = 1 \mid Z = z)$. Let (Z, Z') be a pair of independent draws from the same marginal distribution F_Z. Assume the distribution of Z has positive density with respect to a σ-finite measure, which consists of the counting measure for discrete coordinates and the Lebesgue measure for continuous coordinates.

To quantify the informational content, we first follow the approach in Manski (1988). For a generic pair of coefficients and a nuisance distribution $(b, G_{\epsilon|X}) \in \mathbb{R}^k \otimes \Theta$, define $\xi\left(b, G_{\epsilon|X}\right) \equiv \{z : p(z) \neq \int 1(\epsilon \leq xb - v)\, dG_{\epsilon|x}\}$ and $\tilde{\xi}\left(b, G_{\epsilon|X}\right) \equiv$

$$\Bigg\{(z, z') : (p(z), p(z')) \neq \left(\int 1(\epsilon \leq xb - v)\, dG_{\epsilon|x},\right.$$
$$\left.\int 1\left(\epsilon \leq x'b - v'\right) dG_{\epsilon|x'}\right)\Bigg\}. \tag{2}$$

In words, the set $\xi(b, G_{\epsilon|X})$ consists of states for which the conditional choice probabilities implied by $(b, G_{\epsilon|X})$ differ from those in the actual data-generating process (DGP) $(\beta, F_{\epsilon|X})$. In comparison, the set $\tilde{\xi}$ in (2) consists of paired states where implied conditional choice probabilities differ from those in the actual DGP. We say β is *identified relative to* $b \neq \beta$ if

$$\text{either } \int 1\{z \in \xi(b, G_{\epsilon|X})\} dF_Z > 0 \quad \text{or}$$

[1] Khan and Tamer (2010) showed in a stylized model that there is no informative partial identification result for the intercept in this case.