EMPIRICAL LIKELIHOOD INFERENCE FOR OAXACA-BLINDER DECOMPOSITION

TAISUKE OTSU AND SHIORI TANAKA

Abstract. This paper proposes an empirical likelihood inference method for the Oaxaca-Blinder decompositions. In contrast to the conventional Wald statistic using the delta method, our approach circumvents the linearization errors and estimation of the variance terms. Furthermore, the shape of the resulting empirical likelihood confidence set is determined flexibly by data. Simulation results illustrate usefulness of the proposed inference method.

1. Introduction

Since the seminal works by Oaxaca (1973) and Blinder (1973), the Oaxaca-Blinder (hereafter, OB) decomposition has been widely used in empirical economic analyses to investigate differences in mean outcomes between groups (see, For tin, Lemieux and Firpo, 2011, for an overview). The OB decompositions are typically obtained as products of (differences of) the OLS estimators and averages of regressors. Statistical inference on the OB decompositions is usually conducted based on their standard errors, which are derived by the delta method; see, Oaxaca and Ransom (1998) and Jann (2008) for the cases of non-stochastic and stochastic regressors, respectively.

To fix the idea, consider a typical form of the OB decomposition $\hat{\theta} = \bar{X}_B'(\hat{\beta}_A - \hat{\beta}_B)$, where $\hat{\beta}_A$ and $\hat{\beta}_B$ are the OLS estimators of the groups A and B, respectively, and $\bar{X}_B$ is the sample mean of the regressors of the group B. As shown in Jann (2008), the standard error of $\hat{\theta}$ is obtained by applying the delta method and estimating the variance components $\mathrm{Var}(\bar{X}_B)$, $\mathrm{Var}(\hat{\beta}_A)$, and $\mathrm{Var}(\hat{\beta}_B)$. As is well known in the econometrics literature (e.g., Gregory and Veal, 1985), the t or Wald test based on the delta method is not invariant to formulations of nonlinear parameter hypotheses and the approximation errors by the delta method may not be negligible in finite samples. Intuitively, even though $\bar{X}_B$ and $\hat{\beta}_A - \hat{\beta}_B$ are well approximated by normal distributions, their product $\hat{\theta}$ may be far from normal in finite samples.

To address this issue on the conventional inference method, we adapt the method of empirical likelihood by Owen (1988) to conduct inference on the OB decomposition parameters. In particular, we construct the empirical likelihood functions of two groups that involve their regression parameters, and then combine these likelihood functions to impose the null hypothesis on the OB decomposition parameters of interest. Our empirical likelihood ratio statistic converges to the chi-squared distribution at the true parameter value so that we can conduct hypothesis testing and interval estimation by using the chi-squared critical values. Although our empirical likelihood inference is computationally more expensive than the conventional inference based on the t or Wald statistic, there are several advantages for our method. First, the proposed empirical likelihood inference does not rely on the delta method so that it is free from the linearization error and invariant to formulations of nonlinear hypotheses. Second, since the empirical likelihood ratio is asymptotically pivotal, there is no need to estimate the variance components, such as $\text{Var}(\bar{X}_B)$, $\text{Var}(\hat{\beta}_A)$, and $\text{Var}(\hat{\beta}_B)$ for $\hat{\theta}$. Third, the empirical likelihood ratio-based confidence set is range preserving and transformation respecting. Fourth, in contrast to the Wald-type confidence set whose shape is constrained to be symmetric around the point estimate, the shape of the empirical likelihood ratio-based confidence set is naturally determined by data. Finally, the empirical likelihood inference can easily accommodate side information on data, such as knowledge on some moments or additional instruments.

To illustrate advantages of the proposed method, we conduct a simulation study to investigate the size and power properties. Our simulation results show that our empirical likelihood test for the OB decomposition parameters shows similar size properties as the conventional t test across different setups. On the other hand, the proposed empirical likelihood and conventional t tests exhibit rather different power properties. In particular, when the error term follows an asymmetric distribution such as log-normal, then the empirical likelihood test shows higher power for the negative side which contains zero and is economically more important. This result illustrates that our empirical likelihood inference can be a useful complement to the conventional one.

2. Main Result

Let $\{Y_{Ai}, X_{Ai}\}_{i=1}^{n_A}$ and $\{Y_{Bi}, X_{Bi}\}_{i=1}^{n_B}$ be two independent and identically distributed (iid) samples for the groups A and B, where Y_A and Y_B are scalar outcome variables and X_A and X_B are k-dimensional regressors. For each group, consider the regression model

$$Y_\ell = X_\ell' \beta_\ell + \epsilon_\ell,$$

for $\ell = A, B$, where β_ℓ is a k-dimensional vector of parameters and the error term ϵ_ℓ is independent of X_ℓ and has zero mean. One common way to decompose the difference of the means of the group outcomes is

$$E[Y_A] - E[Y_B] = (E[X_A] - E[X_B])' \beta_A + E[X_B]'(\beta_A - \beta_B),$$

where the first term is the outcome difference explained by the difference in the regressors, and the second term is the unexplained part (often attributed to discrimination). Hereafter, we focus on inference for the second term $\theta = E[X_B]'(\beta_A - \beta_B)$. However, our empirical likelihood approach can be easily extended to conduct inference on the first term or other decompositions as far as they are expressed by (nonlinear) functions of regression parameters and means of regressors.

Let $\bar{X}_\ell$ be the sample mean of X_ℓ, and $\hat{\beta}_\ell$ be the OLS estimator for the regression of Y_ℓ on X_ℓ. Then the point estimator of θ is given by $\hat{\theta} = \bar{X}_B'(\hat{\beta}_A - \hat{\beta}_B)$. By applying the delta method, Jann (2008) suggested to construct the t-ratio by estimating the asymptotic variance of $\hat{\theta}$ as

$$\widehat{\text{Var}}(\hat{\theta}) = \bar{X}_B' \{\widehat{\text{Var}}(\hat{\beta}_A) + \widehat{\text{Var}}(\hat{\beta}_B)\} \bar{X}_B + (\hat{\beta}_A - \hat{\beta}_B)' \widehat{\text{Var}}(\bar{X}_B)(\hat{\beta}_A - \hat{\beta}_B).$$

We note that this variance estimator is derived via linearization of $\hat{\theta} = \bar{X}_B'(\hat{\beta}_A - \hat{\beta}_B)$ around $(\bar{X}_B, \hat{\beta}_A, \hat{\beta}_B) = (E[X_B], \beta_A, \beta_B)$ and requires the variance estimators $\widehat{\text{Var}}(\hat{\beta}_A)$, $\widehat{\text{Var}}(\hat{\beta}_B)$, and $\widehat{\text{Var}}(\bar{X}_B)$.

This paper proposes an alternative inference method for $\theta = E[X_B]'(\beta_A - \beta_B)$ based on empirical likelihood. For the sample $\{Y_{Ai}, X_{Ai}\}_{i=1}^{n_A}$ of the group A, by using the moment condition $E[X_A(Y_A - X_A'\beta_A)] = 0$, the empirical likelihood function for β_A is obtained as

$$\ell_A(\beta_A) = -2 \max_{\{p_{Ai}\}_{i=1}^{n_A}} \sum_{i=1}^{n_A} \log(n_A p_{Ai}), \qquad \text{s.t. } p_{Ai} \geq 0, \quad \sum_{i=1}^{n_A} p_{Ai} = 1, \quad \sum_{i=1}^{n_A} p_{Ai} X_{Ai}(Y_{Ai} - X_{Ai}'\beta_A) = 0.$$

By the Lagrange multiplier method, its dual form is written as

$$\ell_A(\beta_A) = 2 \sup_{\lambda_A} \sum_{i=1}^{n_A} \log(1 + \lambda_A' X_{Ai}(Y_{Ai} - X_{Ai}'\beta_A)),$$

and we employ this dual form to compute $\ell_A(\beta_A)$ in practice. We next construct the empirical likelihood function for the sample $\{Y_{Bi}, X_{Bi}\}_{i=1}^{n_B}$ of the group B. Since θ involves β_B and $\mu_B = E[X_B]$, the empirical likelihood function for the group B is constructed as

$$\ell_B(\beta_B, \mu_B) = -2 \max_{\{p_{Bj}\}_{j=1}^{n_B}} \sum_{j=1}^{n_B} \log(n_B p_{Bj}), \qquad p_{Bj} \geq 0, \quad \sum_{j=1}^{n_B} p_{Bj} = 1, \quad \sum_{j=1}^{n_B} p_{Bj} \begin{bmatrix} X_{Bi}(Y_{Bi} - X_{Bi}'\beta_B) \\ X_{Bi} - \mu_B \end{bmatrix} = 0,$$

and its dual form is

$$\ell_B(\beta_B, \mu_B) = 2 \sup_{\lambda_B} \sum_{j=1}^{n_B} \log\left(1 + \lambda_B' \begin{bmatrix} X_{Bi}(Y_{Bi} - X_{Bi}'\beta_B) \\ X_{Bi} - \mu_B \end{bmatrix}\right).$$

By combining the empirical likelihood functions $\ell_A(\beta_A)$ and $\ell_B(\beta_B, \mu_B)$, we obtain the (profile) empirical likelihood function for θ as

$$L(\theta) = \min_{(\beta_A, \beta_B, \mu_B): \theta = \mu_B'(\beta_A - \beta_B)} \{\ell_A(\beta_A) + \ell_B(\beta_B, \mu_B)\}.$$

The asymptotic property of $L(\theta)$ is obtained as follows.

Proposition. *Suppose $\{Y_{Ai}, X_{Ai}\}_{i=1}^{n_A}$ and $\{Y_{Bi}, X_{Bi}\}_{i=1}^{n_B}$ are independent and identically distributed (iid) sequences with finite fourth moments. Also, $\{Y_{Ai}, X_{Ai}\}_{i=1}^{n_A}$ and $\{Y_{Bi}, X_{Bi}\}_{i=1}^{n_B}$ are independent. Then*

$$L(\theta) \xrightarrow{d} \chi_1^2,$$

as $n_A, n_B \to \infty$ with $n_A/n_B \to c \in (0, \infty)$.

This proposition can be shown as follows. First, by adapting the proof of Jing (1995, Theorem 2) (i.e., replacing "$X_i - \mu_x$" and "$Y_i - \mu_y$" in Jing (1995) with "$X_{Ai}(Y_{Ai} - X_{Ai}'\beta_A)$" and "$\begin{bmatrix} X_{Bi}(Y_{Bi} - X_{Bi}'\beta_B) \\ X_{Bi} - \mu_B \end{bmatrix}$", respectively), we obtain $\ell_A(\beta_A) + \ell_B(\beta_B, \mu_B) \xrightarrow{d} \chi_{3k}^2$. Second, since

$\theta = \mu_B'(\beta_A - \beta_B)$ is a smooth function of $(\beta_A, \beta_B, \mu_B)$, the argument in Hall and LaScala (1990, Theorem 2.1) applied to $\ell_A(\beta_A) + \ell_B(\beta_B, \mu_B)$ yields the conclusion.

This proposition shows that our empirical likelihood statistic $L(\theta)$ achieves asymptotic pivotalness without estimating the variance components such as $\text{Var}(\bar{X}_B)$, $\text{Var}(\hat{\beta}_A)$, and $\text{Var}(\hat{\beta}_B)$. Based on this proposition, the $100(1-\alpha)\%$ empirical likelihood confidence set of θ is given by $ECLI_\theta = \{\theta : L(\theta) \leq \chi^2_{1,\alpha}\}$, where $\chi^2_{1,\alpha}$ is the $(1-\alpha)$-th quantile of the χ^2_1 distribution.

Although we focus on the unexplained (discrimination) part of decomposition, our method can be applied to the other term of the twofold decomposition $(\bar{X}_A - \bar{X}_B)'\hat{\beta}_A$ (explained or composition effect). In addition, as Jann (2008) explained, there is a threefold decomposition where the overall difference is decomposed of three components including an interaction term between the difference of means and that of coefficients. Indeed, the twofold decomposition is the equivalent conversion of the threefold one with introducing the benchmark coefficient ($\hat{\beta}_A$ here). Again, we can build an analogous inference method for the threefold decomposition with the empirical likelihood approach.

3. Simulation

We conduct a simulation study to check the size and power properties of our empirical likelihood method in comparison to Jann's (2008) t ratio and the delta method. As a simulation setup, we consider

$$Y_{Ai} = 7X_{Ai} + \epsilon_{Ai}, \qquad Y_{Bi} = 5X_{Bi} + \epsilon_{Bi},$$

where $\epsilon_{gi} \sim N(0,1)$ and X_{gi} and ϵ_{gi} are independent for $g = A, B$. For X_{Ai} and X_{Bi}, we consider the cases of normal, χ^2_3, and log-normal with mean 1 and variance 1. We also consider heteroskedastic error terms $\epsilon_{gi} = \sqrt{0.5 + 0.5X^2_{gi}}v_{gi}$ with $v_{gi} \sim N(0,1)$. We are interested the OB decomposition parameter $\theta = \mu_B'(\beta_A - \beta_B) = 2$. The sample size is set as $n = 100$. The number of Monte Carlo replications is 500 for both the size and power analyses.

We first compare the size of three methods: Jann's t ratio (Jann), delta method (Delta), and empirical likelihood (EL). The delta method included into comparison since previous research such as Oaxaca and Ransom (1998) and Greene (2008) employs it to estimate the variance of the decomposition estimators. The rejection frequencies at 5% significance level of the three methods are presented in Table 1. There are no significant differences in the size properties among these methods. This is true even for asymmetric distributions, such as χ^2_3 and Log-normal, and heteroskedasticity.

Table 1. Simulation result: Size

	Jann	Delta	EL
Normal	0.06	0.06	0.062
χ^2_3	0.064	0.064	0.064
Log-normal	0.086	0.086	0.072
Normal heteroskedastic	0.054	0.054	0.05
χ^2_3 heteroskedastic	0.066	0.066	0.068
Log-normal heteroskedastic	0.11	0.11	0.096

We next compare the power of these methods by setting the null hypothesis $H_0 : \theta = 2 + c$ with $c \in \{-2.0, -1.9, \ldots, 1.9, 2.0\}$. Since we are interested in the performance of our method (EL) in the cases where the data are asymmetrically distributed, we present the power curves for the cases of normal heteroskedastic and log-normal heteroskedastic errors in Figure 1. Although there is no significant differences in the normal heteroskedastic case, EL performs better on the negative side of the deviation in the log-normal heteroskedastic case. To illustrate this point, we draw the empirical likelihood function $L(\theta)$ in Figure 2, where the horizontal line is the .95-th quantile of the χ^2_1 distribution used for the critical value. Due to the positive skewness of $L(\theta)$, the rejection rate is higher on the left side of the true value $\theta = 2$. We note that the left side contains zero (i.e., no discrimination in the OB decomposition parameter). Thus in this example, EL exhibits better power properties than other methods to detect presence of discrimination. Overall EL can be a useful complement to the conventional t ratio in this context.

FIGURE 1. Simulation result: Power

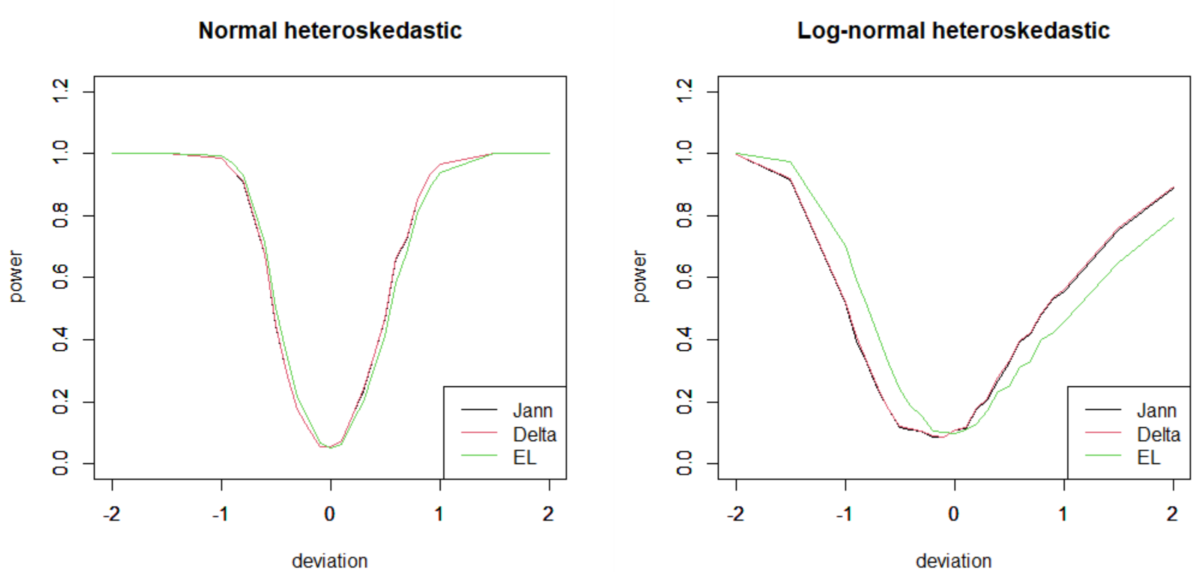
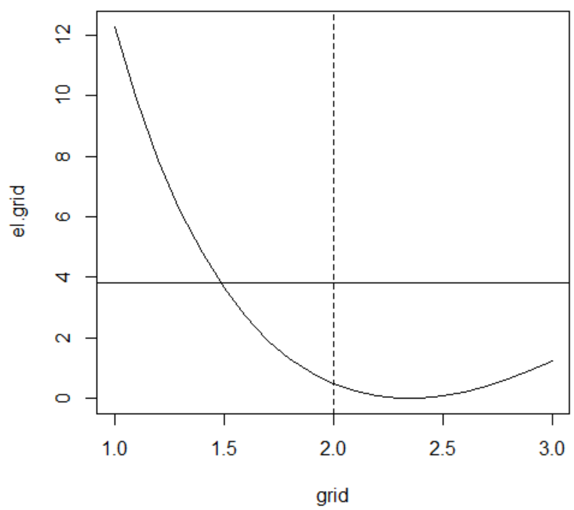


Figure 2. Statistical Inference



References

[1] Blinder, A. S. (1973) Wage discrimination: reduced form and structural estimates, *Journal of Human Resources*, 8, 436-455.

[2] Fortin, N., T. Lemieux and S. Firpo (2011) Decomposition methods in economics, in Ashenfelter, O. and D. Card (eds.), *Handbook of Labor Economics*, 4a, 1-102, North Holland, Amsterdam.

[3] Gregory, A. W. and M. R. Veall (1985) Formulating Wald tests of nonlinear restrictions, *Econometrica*, 53, 1465-1468.

[4] Greene, W. H. (2008) Econometric Analysis, 6th ed., Upper Saddle River, New Jersey.

[5] Hall, P. and B. La Scala (1990) Methodology and Algorithms of Empirical Likelihood, *International Statistical Review*, 58(2), 109-127.

[6] Jann, B. (2008) The Blinder–Oaxaca decomposition for linear regression models, *Stata Journal*, 8, 453-479.

[7] Jing, B. (1995) Two-sample empirical likelihood method, *Statistics & Probability Letters*, 24, 315-319.

[8] Oaxaca, R. L. (1973) Male-Female wage differentials in urban labor markets, *International Economic Review*, 14, 693-709.

[9] Oaxaca, R. L. and L. Ransom (1998) Calculation of approximate variances for wage decomposition differentials, *Journal of Economic and Social Measurement*, 24, 55-61.

[10] Owen, A. (1988) Empirical likelihood ratio confidence intervals for a single functional, *Biometrika*, 75, 237-249.

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

Email address: `t.otsu@lse.ac.uk`

Graduate School of Economics, Keio University, 2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan.

Email address: `tshio1997@keio.jp`