Available online at www.sciencedirect.com

SCIENCE DIRECT®

Economics Letters 86 (2005) 413–417

economics letters

www.elsevier.com/locate/econbase

A computational trick for delta-method standard errors

Leslie E. Papke, Jeffrey M. Wooldridge*

Department of Economics, Michigan State University, East Lansing, MI 48824-1038, United States

Received 30 March 2004; accepted 23 July 2004
Available online 7 December 2004

Abstract

We show how to compute the standard error for a nonlinear function of regression coefficients using a simple substitution trick. We use the method to obtain a standard error for the long-run effect in a dynamic panel data model.

Keywords: Asymptotic standard error; Delta method; Nonlinear function; Robust standard error

JEL classification: C10; C63

1. Introduction

Sometimes in empirical studies, one needs to compute the asymptotic standard error for a nonlinear function of underlying model parameters. For example, the long-run effect in an infinite distributed lag model is a nonlinear function of the coefficients on the explanatory variables and the lagged dependent variable. While bootstrap methods—see, for example, Horowitz (2001)—are being used more often for obtaining standard errors, the delta method is still important because it is widely applicable and does not require potentially costly resampling. Wooldridge (2002, Section 3.5.2) contains a general discussion of the delta method.

One hurdle in applying the delta method, even for a single nonlinear function of parameters, is that fairly sophisticated software has been needed to compute the standard error (Stata 8 now has a

* Corresponding author. Tel.: +1 517 353 5972; fax: +1 517 432 1068.
E-mail address: wooldri1@msu.edu (J.M. Wooldridge).

0165-1765/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.econlet.2004.07.022

command, "nlcom," that can be used to compute a delta-method standard error). Unfortunately, many researchers worldwide do not have access to expensive software that reports delta-method standard errors. Consequently, standard errors for nonlinear functions are computed less often than they should be.

The computational trick we propose in this note extends a method used to obtain a standard error for linear combinations of regression parameters; see, for example, Wooldridge (2003, Section 4.4 and Appendix E). Because our method relies on estimating a linear transformation of the original model, a delta-method standard error can be obtained using software that supports only basic regression analysis. In Section 4, we apply our method to obtain a standard error for the long-run effects of school spending in a dynamic panel data model.

2. Computing the standard error for a linear combination

We start with the standard linear model $y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_k x_{tk} + u_t = \mathbf{x}_t \boldsymbol{\beta} + u_t$, $t = 1, \ldots, n$, where $\mathbf{x}_t$ is $1 \times K$ and $\boldsymbol{\beta}$ is $K \times 1$; typically, $x_t 1 \equiv 1$. In matrix notation,

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{u}, \tag{1}$$

where $\boldsymbol{y}$ is an $n \times 1$ vector of observations on the y_t, $\mathbf{X}$ is the $n \times k$ matrix of explanatory variables, and $\boldsymbol{u}$ is an $n \times 1$ vector of errors. As this paper is purely about computation, model assumptions are not needed, although our method of obtaining $\hat{\boldsymbol{\beta}}$ and computing its estimated asymptotic variance matrix are always motivated by different sets of assumptions.

Initially, let $\hat{\boldsymbol{\beta}}$ be the ordinary least squares (OLS) estimate of $\boldsymbol{\beta}$. Then, under asymptotic versions of the Gauss–Markov assumptions—see Wooldridge (2003, Chapter 11)—the appropriate asymptotic variance matrix estimator is

$$\hat{\sigma}^2 (\mathbf{X}' \mathbf{X})^{-1}, \tag{2}$$

where $\hat{\sigma}^2 = (n-k)^{-1} \sum_{t=1}^{n} \hat{u}_t^2$ and the $\hat{u}_t$ are the OLS residuals. A heteroskedasticity-robust version is

$$(\mathbf{X}' \mathbf{X})^{-1} \left(\sum_{t=1}^{n} \hat{u}_t^2 \boldsymbol{x}_t' \boldsymbol{x}_t \right) (\mathbf{X}' \mathbf{X})^{-1}, \tag{3}$$

where $\boldsymbol{x}_t$ is the t^{th} row of $\mathbf{X}$. Formulas robust to general serial correlation as well as heteroskedasticity are also available, but we will not need those formulas explicitly.

Now, suppose that $\boldsymbol{a}$ is a $1 \times k$ vector of constants, and we are interested in the linear combination $\theta \equiv \boldsymbol{a}\boldsymbol{\beta} = a_1 \beta_1 + a_2 \beta_2 + \ldots + a_k \beta_k$. If $\hat{\mathbf{V}} = \widehat{\text{Avar}}(\hat{\boldsymbol{\beta}})$ then $\widehat{\text{Avar}}(\hat{\theta}) = \boldsymbol{a}\hat{\mathbf{V}}\boldsymbol{a}'$, and so the asymptotic standard error of $\hat{\theta}$ is

$$se\left(\hat{\theta}\right) = \sqrt{\boldsymbol{a}\hat{\mathbf{V}}\boldsymbol{a}'}. \tag{4}$$

As shown in Wooldridge (2003, Appendix E), when $\hat{\mathbf{V}}$ has the form (2), se$(\hat{\theta})$ is easily obtained as follows. For some j with $a_j \neq 0$, write β_j in terms of $\boldsymbol{a}$, θ, and other elements of $\boldsymbol{\beta}$. Then, substitute into the original model so that the parameters in the new model are $\beta_1,\ldots,\beta_{j-1},\theta,\beta_{j+1},\ldots,\beta_k$. More precisely, the new OLS regression is

$$\begin{aligned} y_t \text{ on } &\left[x_{t1}-(a_1/a_j)x_{tj}\right],\ldots,\left[x_{t,j-1}-(a_{j-1}/a_j)x_{tj}\right],x_{tj}/a_j,\\ &\left[x_{t,j+1}-(a_{j+1}/a_j)x_{tj}\right],\ldots,\left[x_{tk}-(a_k/a_j)x_{tj}\right],t=1,\ldots n, \end{aligned} \tag{5}$$

and the coefficient on x_{tj}/a_j is $\hat{\theta}$. More importantly, the reported standard error on x_{tj}/a_j is exactly as Eq. (4).

That the same trick works for robust versions of $\hat{\mathbf{V}}$ follows because OLS residuals are invariant to regression on nonsingular linear transformations of the regressors. So, for example, if $\mathbf{G}$ is a $k\times k$ nonsingular matrix, the regression $\boldsymbol{y}$ on $\mathbf{XG}$ gives an estimated variance matrix $\mathbf{G}^{-1}\hat{\mathbf{V}}\mathbf{G}'^{\,-1}$ for any of the robust estimates $\hat{\mathbf{V}}$ used for regression analysis, whether it is the White (1980) heteroskedasticity-robust variance matrix estimator in Eq. (3) or the Newey and West (1987) estimator that is also robust to serial correlation of unknown form.

Since residuals in instrumental variables (IV) contexts are invariant to nonsingular linear transformations of the explanatory variables, and the forms of the asymptotic variances are such that IV estimation with $\mathbf{XG}$ as the explanatory variables gives a variance matrix $\mathbf{G}^{-1}\hat{\mathbf{V}}\mathbf{G}'^{\,-1}$, the same substitution trick still works. For example, the usual two-stage least squares (2SLS) variance matrix estimator is

$$\hat{\sigma}^2\left[\mathbf{X}'\,\mathbf{Z}(\mathbf{Z}'\,\mathbf{Z})^{-1}\mathbf{Z}'\,\mathbf{X}\right]^{-1}; \tag{6}$$

if we replace $\mathbf{X}$ with $\mathbf{XG}$, we get $\mathbf{G}^{-1}\hat{\mathbf{V}}\mathbf{G}'^{\,-1}$, with $\hat{\mathbf{V}}$ as in Eq. (6). The same holds for robust variance matrices for 2SLS; see Wooldridge (2002, Eq. (5.34)).

Not surprisingly, all of the previous comments extend to more complicated data structures and estimation methods. In estimating a linear model using panel data, we can use pooled OLS, fixed effects, random effects, and instrumental variables versions of all three. All variance matrix estimators, even fully robust ones, have structures that imply the basic substitution method produces the appropriate standard error (for example, see Eqs. (7.26) and (10.59) in Wooldridge, 2002)

3. A delta-method standard error

We now show how the simple trick outlined in Section 2 can be adapted to obtain a delta-method standard error for a nonlinear function. Let r: $\mathbb{R}^k\rightarrow\mathbb{R}$ be a continuously differentiable function, and define $\theta\equiv r(\boldsymbol{\beta})$. Let $\mathbf{g}(\boldsymbol{\beta})$ be the $1\times k$ gradient of $r(\bullet)$. If $\hat{\boldsymbol{\beta}}$ is the estimator of $\boldsymbol{\beta}$ and $\hat{\mathbf{V}}$ is its asymptotic variance, the asymptotic standard deviation of $\hat{\theta}\equiv r(\hat{\boldsymbol{\beta}})$

$$se\left(\hat{\theta}\right)=\sqrt{\hat{\mathbf{g}}\hat{\mathbf{V}}\hat{\mathbf{g}}'}, \tag{7}$$

where $\hat{\mathbf{g}}\equiv\mathbf{g}(\hat{\boldsymbol{\beta}})=(\hat{g}_1,\ldots,\hat{g}_k)$ is the gradient evaluated at $\hat{\boldsymbol{\beta}}$. Not surprisingly, because the delta method works off of linearizing the nonlinear function, Eq. (7) has the same form as Eq. (4) with $\boldsymbol{a}\equiv\hat{\mathbf{g}}$. This means we can use the same substitution trick to obtain Eq. (7). One important difference is that unlike in the linear case, estimation using transformed regressors does not produce $\hat{\theta}$. But $\hat{\theta}$ is easily gotten by plugging $\hat{\boldsymbol{\beta}}$ into $r(\bullet)$. The extra work is in obtaining se($\hat{\theta}$), and we can use the substitution method to simplify the calculations. To obtain the transformed explanatory variables, define a pseudoparameter as

$\delta = \hat{\mathbf{g}}\boldsymbol{\beta} = \hat{g}_1\beta_1 + \hat{g}_2\beta_2 + \ldots + \hat{g}_k\beta_k$. We do not have any interest in estimating δ. Nevertheless, if we act as if δ is the parameter of interest and ignore that the $\hat{g}_j$ is estimated, then the same substitution trick used in regression (5) produces Eq. (7), the asymptotic standard error of $\hat{\theta}$ (effectively, the asymptotic standard errors of $\hat{\theta}$ and $\delta \equiv \hat{g}_1\hat{\beta}_1 + \hat{g}_2\hat{\beta}_2 + \ldots + \hat{g}_k\hat{\beta}_k$ are the same where estimation error in the $\hat{g}_h$ is ignored). We summarize the steps:

(1) Define the parameter of interest as $\theta = r(\boldsymbol{\beta})$ and find the gradient $\mathbf{g}(\boldsymbol{\beta})$ of $r(\boldsymbol{\beta})$. At most, this requires computing the k partial derivatives $\partial r(\boldsymbol{\beta})/\partial \beta_h$, $h=1,\ldots, k$.
(2) Evaluate $\mathbf{g}(\boldsymbol{\beta})$ at $\hat{\boldsymbol{\beta}}$ to obtain $\hat{\mathbf{g}} \equiv (\hat{g}_1, \hat{g}_2 \ldots, \hat{g}_k$; entries that are zero can be dropped to form a condensed $\hat{\mathbf{g}}$). $\hat{\boldsymbol{\beta}}$ can be any of the estimators described in Section 2.
(3) Choose a nonzero element of $\hat{\mathbf{g}}$, say $\hat{g}_j$. Define transformed regressors as $\tilde{x}_h \equiv [x_h - (\hat{g}_h/\hat{g}_j)x_j]$, $h \neq j$ and $\tilde{x}_j \equiv x_j/\hat{g}_j$, where the transformation is for each observation. If $r(\boldsymbol{\beta})$ does not depend on β_h, then $\hat{g}_h = 0$, and so $\tilde{x}_h = x_h$.
(4) Estimate an equation—by OLS, IV, or panel data versions of these, including fixed effects and random effects—with the $\tilde{x}_h$, $h=1,\ldots, k$, as explanatory variables, where any IVs are the same as in the original estimation. The standard error for the coefficient on $\tilde{x}_j$ is Eq. (7).

Compared with other strategies, this procedure allows computation of delta-method standard errors using the most rudimentary regression packages. Plus, in more sophisticated packages, robust standard errors can be obtained with a simple change in options.

4. An example: a standard error for the long-run effect in a panel data-distributed lag model

Papke (2001) uses school-level panel data for Michigan to estimate the effects of spending on fourth-grade math test pass rates. In one specification, she includes a lagged-dependent variable, along with current and lagged spending, and estimates the parameters by pooled OLS. It is common in the school production function literature to add lagged scores or pass rates in order to control for historical differences in school performance.

Here we use district-level data for the years 1993 through 1998 (with 1992 being used to obtain the initial lags of the pass rates and spending variables). The model is

$$\begin{aligned} math4_{it} = {} & \alpha_0 + \alpha_1 y94_t + \ldots + \alpha_5 y98_t + \beta_1 math4_{i,t-1} + \beta_2 lrexpp_{it} + \beta_3 lrexpp_{i,t-1} \\ & + \beta_4 lunch_{it} + \beta_5 lunch_{it}^2 + \beta_6 lenroll_{it} + \beta_7 lenroll_{it}^2 + u_{it}, \end{aligned} \tag{8}$$

where i indexes school and t indexes year. The variable math4_{it} is the percent of fourth graders passing the MEAP math test, lrexpp_{it} is the log of real expenditures per pupil, lunch_{it} is the percent of students eligible for the federal free and reduced price lunch program, and lenroll_{it} is log of total school enrollment. The long-run effect of a permanent increase in spending is $\theta = (1-\beta_1)^{-1}(\beta_2+\beta_3)$; more precisely, $\theta/10$ is the eventual percentage-point change in math4, given a permanent 10% increase in spending. The gradient we need to obtain the delta-method standard error, evaluated at the estimates, is

$$\hat{\mathbf{g}} = \left[\left(1-\hat{\beta}_1\right)^{-2}\left(\hat{\beta}_2+\hat{\beta}_3\right), 1/\left(1-\hat{\beta}_1\right), 1/\left(1-\hat{\beta}_1\right)\right]. \tag{9}$$

Using 550 school districts in Michigan, the needed estimates are $\hat{\beta}_1$=0.513, $\hat{\beta}_2$=−0.500, and $\hat{\beta}_3$=5.918. Therefore, the estimated long-run effect is 11.13, which means that a permanent 10% increase in spending increases the estimated pass rate by about 1.1 percentage points. The estimated gradient (with respect to the parameters in the long-run effect) is $\hat{\mathbf{g}}$=(22.844, 2.053, 2.053), and so, we can define $\tilde{x}_{it1}=\mathrm{math4}_{i,t-1}/22.844$, $\tilde{x}_{it2}=\mathrm{lrexpp}_{it}-(2.053/22.844)\mathrm{math4}_{i,t-1}$, and $\tilde{x}_{it3}=\mathrm{lrexpp}_{i,t-1}-(2.053/22.844)\mathrm{math4}_{i,t-1}$. The pooled OLS regression of $\mathrm{math4}_{it}$ on 1, $y94_t$,..., $y98_t$, $\tilde{x}_{it1}$, $\tilde{x}_{it2}$, $\tilde{x}_{it3}$, lunch_{it}, lunch_{it}^2, $\mathrm{lenroll}_{it}$, $\mathrm{lenroll}_{it}^2$ (across all 3300 observations) gives as the standard error on $\tilde{x}_{it1}$ the value 2.35. The heteroskedasticity-robust standard error is 3.23, which leads to a 95% confidence interval for θ of 4.80 to 17.46 (in Stata 8, using the "nlcom" command, the standard error is given, to three decimal places, as 3.229).

The same trick works if we add a district fixed effect to Eq. (8), difference to remove the fixed effect, and use instrumental variables for $\Delta\mathrm{math4}_{i,t-1}$ in the first-differenced equation, as in Arellano and Bond (1991). Here, we use $\mathrm{math4}_{i,t-2}$ as a single IV for $\Delta\mathrm{math4}_{i,t-1}$ (with all other differenced explanatory variables acting as their own IVs) and use pooled 2SLS, being sure to compute a standard error robust to heteroskedasticity and serial correlation because differencing induces serial correlation in the idiosyncratic error. The estimates are $\hat{\beta}_1$=0.0954, $\hat{\beta}_2$=−0.893, and $\hat{\beta}_3$=11.377. While the coefficient on the lagged pass rate is now much smaller, the coefficient on lagged spending is notably larger. The estimated long-run effect is now 11.59, which is slightly higher than before but pretty close to the estimate without a district fixed effect. Using Eq. (9), the estimated gradient is now $\hat{\mathbf{g}}$=(12.812, 1.105, 1.105). So, let $\Delta\tilde{x}_{it1}\equiv\Delta\mathrm{math4}_{i,t-1}/12.812$, $\Delta\tilde{x}_{it2}\equiv\Delta\mathrm{lrexpp}_{it}-(1.105/12.812)\Delta\mathrm{math4}_{i,t-1}$, and $\Delta\tilde{x}_{it3}\equiv\Delta\mathrm{lrexpp}_{i,t-1}-(1.105/12.812)\Delta\mathrm{math4}_{i,t-1}$. Using these transformed variables in pooled 2SLS along with the other differenced variables, where the instrument list is the same as before, gives a fully robust standard error for $\hat{\theta}$ (as the standard error associated with $\Delta\tilde{x}_{it1}$) equal to 7.25, which means the estimated long-run effect is only marginally significant (t=1.60). Because we have differenced and used IV, this t statistic for $\hat{\theta}$ is much smaller than that from the pooled OLS estimation—although the estimated long-run effect is similar.

References

Arellano, M., Bond, S.R., 1991. Some specification tests for panel data: Monte Carlo evidence and an application to employment equations. Review of Economic Studies 58, 277–298.

Horowitz, J.L., 2001. The bootstrap in econometrics. In: Leamer, E., Heckman, J.J. (Eds.), Handbook of Econometrics, vol. 5. Amsterdam, North Holland, pp. 3159–3228.

Newey, W.K., West, K.D., 1987. A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Papke, L.E., 2001. The effects of spending on test pass rates: evidence from Michigan, Mimeo. Michigan State University, Department of Economics.

White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817–838.

Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, MA.

Wooldridge, J.M., 2003. Introductory Econometrics: A Modern Approach. South-Western, Cincinnati, OH.