International Journal of Applied Mathematics

Volume 29 No. 3 2016, 395-400
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
doi: http://dx.doi.org/10.12732/ijam.v29i3.11

A SIMPLE HEURISTIC NOTE ON RANDOM SLOPES AND INTERCEPTS COEXISTING IN REGRESSION

Gregory L. Light
Department of Finance
Providence College
Providence, Rhode Island, 02918, USA

Abstract: We present a new treatment on the problem in simple linear regression that has random disturbances in both the intercept and the slope by an ad hoc but heuristic example. Our approach is distinguished by its simplicity and estimation unbiasedness.

AMS Subject Classification: 62J05, 62J12, 62J99
Key Words: regression random coefficients, heteroskedasticity

1. Introduction

Standard simple linear regression models assign the random term to the intercept, as patently revealed in the following basic equation in business economics,

$$\begin{aligned} \text{total cost } TC \;&=\; \text{fixed cost } FC \\ &+\; \text{average variable cost } AVC \cdot \text{ production quantity } q + \epsilon, \end{aligned}$$

which begs the question: What if AVC also carries a random term? Our literature research found that this topic of random coefficients in regression had received general attention from diverse fields. Three generic treatments stood out: Bayesian inferences with simulations (e.g., $[1, 4]$), order statistics with

Received: April 30, 2016

specialized samplings (e.g., [3]), and application of instrumental variables (e.g., [5]). Our treatment here distinguishes itself by its simplicity and estimation unbiasedness and highlights the essence of the problem being "heteroscedasticity on both ends", i.e., as x goes to zero as well as to infinity. We note that the problem here does not reduce to a combination of heteroskedasticity and measurement error in the dependent variable.

2. Analysis

Textbooks on econometrics (e.g., [2], p. 258-261) have treated

$$TC = FC + (AVC + \epsilon_{slope}) \cdot q, \tag{2.1}$$

by

$$AC \equiv \frac{TC}{q} = \frac{FC}{q} + AVC + \epsilon_{slope}, \tag{2.2}$$

as a special case in heteroskedasticity, where FC does not carry a random term. However, our problem on hand is

$$\begin{aligned} \frac{TC}{q} &= \frac{FC + \epsilon_{intercept}}{q} + AVC + \epsilon_{slope}, \quad \text{or} \\ TC &= FC + AVC \cdot q + (\epsilon_{slope} \cdot q + \epsilon_{intercept}) \qquad (2.3) \\ &\equiv FC + AVC \cdot q + \epsilon_{combined}, \qquad (2.4) \end{aligned}$$

where $\epsilon_{combined}$ has a normal distribution with $mean = 0$ and variance

$$Var\left(\epsilon_{combined}\right) = q^2 \sigma^2_{slope} + \sigma^2_{intercept}. \tag{2.5}$$

Clearly as $q \to 0$, $\left(\sigma^2_{intercept}/q\right)$ begins to vary more and more, so that the formulation by (TC/q) suffers from heteroskedasticity just as the formulation by TC does for $q \to \infty$.

Consider now the variable, changes in (TC/q), with $q_2 > q_1$:

$$\begin{aligned} &\frac{TC(q_2)}{q_2} - \frac{TC(q_1)}{q_1} \\ = \ &\left(\frac{FC + \epsilon_{intercept}}{q_2} + AVC + \epsilon_{slope}\right) \\ &-\left(\frac{FC + \epsilon_{intercept}}{q_1} + AVC + \epsilon_{slope}\right) \\ = \ &FC \cdot \left(\frac{q_1 - q_2}{q_2 q_1}\right) + \epsilon_{\triangle AC}, \qquad (2.6) \end{aligned}$$

where $\epsilon_{\triangle AC}$ has a normal distribution with $mean = 0$ and variance

$$Var\left(\epsilon_{\triangle AC}\right) = \sigma^2_{intercept}\left(\frac{1}{q_2^2}+\frac{1}{q_1^2}\right)+2\sigma^2_{slope}, \tag{2.7}$$

that is,

$$\frac{TC\left(q_2\right)}{q_2}-\frac{TC\left(q_1\right)}{q_1}\equiv X \tag{2.8}$$

is a random variable with

$$\begin{aligned} E\left(X\right) &= FC\cdot\left(\frac{q_1-q_2}{q_2 q_1}\right) \text{ and} && (2.9)\\ Var\left(X\right) &= Var\left(\epsilon_{\triangle AC}\right), && (2.10)\end{aligned}$$

suggesting forgoing regression altogether and applying a simple estimation of $E\left(X\right)$ by the sample-mean $\bar{x}$ on a random sample of size n (which is known to be unbiased among other desirable estimation properties), as illustrated below.

Example 1. Consider

$$\begin{aligned} TC &= 100+10q+\epsilon_{slope}q+\epsilon_{intercept}, \text{ with}\\ Var\left(\epsilon_{slope}\right) &= 1 \text{ and } Var\left(\epsilon_{intercept}\right)=16,\\ \text{for q} &\in \left\{9,10\right\},\ n=100, \text{ for}\\ X &\equiv \frac{TC\left(10\right)}{10}-\frac{TC\left(9\right)}{9}. && (2.11)\end{aligned}$$

Then,

$$\begin{aligned} E\left(X\right) &= -100\times\frac{1}{90} \text{ or}\\ FC &= -90E\left(X\right), \text{ and}\\ Var\left(X\right) &= 16\left(\frac{1}{100}+\frac{1}{81}\right)+2\approx 2.36, && (2.12)\end{aligned}$$

so that $90\sqrt{2.36}\approx 138$ gives the standard deviation of estimating FC by subtracting the average cost at $q=9$ from that at $q=10$ for $n=1$ observation. Thus, for $n=100$ we have

$$\begin{aligned} E\left(-90\bar{x}\right) &= 100 \text{ with}\\ \text{the true standard error} &= \frac{138}{\sqrt{100}}=13.8. && (2.13)\end{aligned}$$

A simulation run yielded

$$\bar{x} = -1.2 \tag{2.14}$$

so that the estimated FC was

$$FC_{estimated} = -1.2 \times (-90) = 108 \tag{2.15}$$

with an estimated standard error of

$$\left(s/\sqrt{100}\right) \approx 14. \tag{2.16}$$

For an estimation of AVC, we deducted $FC_{estimated} = 108$ from the observed TC_i, $i = 1, 2, \cdots, 100$, and divided the estimated variable cost by the production quantity $q = 10$ to arrive at the estimated AVC_i and

$$\begin{aligned} & \frac{1}{100}\sum_{i=1}^{100} AVC_{i,estimated} && (2.17)\\ = \;& AVC_{estimated} \\ = \;& \frac{1}{100}\sum_{i=1}^{100}\left(\frac{TC_i\,(10) - 108}{10}\right) = 9.2, && (2.18)\\ & \text{with an estimated standard error} && (2.19)\\ = \;& \sqrt{1.17 + 1.4} \approx 1.6, \end{aligned}$$

where 1.17 = the sample-variance of $\left\{\frac{TC_i(10)}{10} \mid i = 1, 2, \cdots, 100\right\}$, and 1.4 = (the estimated standard error 14 from Equation (2.16) in the estimation of FC by 108)/$(q = 10)$. Finally we estimated $Var\,(\epsilon_{slope})$ and $Var\,(\epsilon_{intercept})$ by solving the two simultaneous equations,

$$\begin{aligned} Var\left(\frac{TC}{10}\right) &= Var\,(\epsilon_{slope}) + Var\left(\frac{1}{10}\epsilon_{intercept}\right), && (2.20)\\ Var\left(\frac{TC}{10} - \frac{TC}{9}\right) &= 2Var\,(\epsilon_{slope}) + Var\left(\frac{1}{10}\epsilon_{intercept}\right) && (2.21)\\ &\quad + Var\left(\frac{1}{9}\epsilon_{intercept}\right), && (2.22) \end{aligned}$$

with

$$\begin{aligned} Var\left(\frac{TC}{10}\right)_{estimated} &= 1.17 \text{ and} && (2.23)\\ Var\left(\frac{TC}{10} - \frac{TC}{9}\right)_{estimated} &= 2.41, && (2.24) \end{aligned}$$

so that the estimated standard deviations

$$\begin{aligned}\hat{\sigma}\left(\epsilon_{slope}\right) &= 0.93, \text{ and} &(2.25)\\ \hat{\sigma}\left(\epsilon_{intercept}\right) &= 5.52. &(2.26)\end{aligned}$$

To summarize, our estimated equation was

$$\begin{aligned}predicted\ TC &= \underset{(14)}{108} + \underset{(1.6)}{9.2}\,q \ \ (= 200 \text{ for } q = 10)\,, &(2.27)\\ &\quad \text{with a standard error}\\ &= \sqrt{5.52^2 + 0.93^2 q^2}\\ &= \sqrt{30.47 + 0.86 q^2}\ \ (= 10.79 \text{ for } q = 10)\,,\end{aligned}$$

the true population equation, to repeat, is

$$\begin{aligned}TC &= 100 + 10q + \epsilon_{slope} q + \epsilon_{intercept} &(2.28)\\ &\quad (= 200 + 10\epsilon_{slope} + \epsilon_{intercept} \text{ for } q = 10)\,,\\ \text{with } \sigma\left(\epsilon_{slope}\right) &= 1 \text{ and } \sigma\left(\epsilon_{intercept}\right) = 4 \text{ (so that}\\ \sigma\left(TC\left(10\right)\right) &= \sqrt{116} \approx 10.77).\end{aligned}$$

To appreciate the effect of $\sigma\left(\epsilon_{intercept}\right)$ on the regression equation

$$\frac{TC}{q} = \frac{FC + \epsilon_{intercept}}{q} + AVC + \epsilon_{slope}$$

for $q \to 0$, we simulated TC for $q = 0.1,\ 0.2, \cdots, 0.9,\ 1,\ 2, \cdots, 10$, and as expected we found that the estimation error could be very large, such as

$$\begin{aligned}predicted\ \frac{TC}{q} &= \frac{103}{q} + 3.7, \text{ or} &(2.29)\\ predicted\ TC &= 103 + 3.7q.\end{aligned}$$

3. Summary

The common exposition of a simple regression equation by $Y_i = \alpha + \beta X_i + \epsilon_i$ tends to deflect one's attention from the fact of $Y_i = (\alpha + \epsilon_i) + \beta X_i$, which naturally prompts the consideration of $Y_i = (\alpha + \epsilon_i) + (\beta + \xi) X_i$. In this paper we presented a solution to this problem via a heuristic example, from which we noted: (1) one might remove those sample observations

$$\{(x_i, y_i) \mid |\sigma(\epsilon_i) - \sigma(\epsilon_{i+1})| >> 0\}$$

and proceed to regress (Y_i/X_i) on $(1/X_i)$ by the ordinary least squares; (2) otherwise, our treatment here could be an alternative.

References

[1] B.A. Coull, A random intercepts–functional slopes model for flexible assessment of susceptibility in longitudinal designs, *Biometrics*, **67**, No 2 (2011), 486-494.

[2] J. Kmenta, *Elements of Econometrics*, Macmillian, New York (1971).

[3] T. Li and N. Balakrishnan, Best linear unbiased estimators of parameters of a simple linear regression model based on ordered ranked set samples, *Journal of Statistical Computation and Simulation*, **78**, No 12 (2008), 1267-1278.

[4] T.R. Ten Have and A.R. Localio, Empirical Bayes estimation of random effects parameters in mixed effects logistic regression models, *Biometrics*, **55**, No 4 (1999), 1022-1029.

[5] Y. Wang, A. Ruan, and Z. Zhan, The numerical simulation of improving parameter estimation by instrumental variable method, *Cybernetes: The International Journal of Systems & Cybernetics*, **41**, No 7-8 (2012), 985-993.