ISSN 1183-1057

Department of Economics

Discussion Papers

02-3

Oh No! I Got The Wrong Sign! What Should I Do?

P. Kennedy

2002

SIMON FRASER UNIVERSITY

Oh No! I Got the Wrong Sign! What Should I Do?

Peter Kennedy
Professor of Economics,
Dept. of Economics
Simon Fraser University
Burnaby, BC
Canada V5A 1S6

Tel. 604-291-4516
Fax: 604-291-5944
Email: kennedy@sfu.ca

Abstract

Getting a “wrong” sign in empirical work is a common phenomenon. Remarkably, econometrics textbooks provide very little information to practitioners on how this problem can arise. This paper exposits a long list of ways in which a “wrong” sign can occur, and how it might be corrected.

Oh No! I Got the Wrong Sign! What Should I Do?

We have all experienced, far too frequently, the frustration caused by finding that the estimated sign on our favorite variable is the opposite of what we anticipated it would be. This is probably the most alarming thing "that gives rise to that almost inevitable disappointment one feels when confronted with a straightforward estimation of one's preferred structural model." (Smith and Brainard, 1976, p.1299). To address this problem, we might naturally seek help from applied econometrics texts, looking for a section entitled "How to deal with the wrong sign." Remarkably, a perusal of existing texts does not turn up sections devoted to this common problem. Most texts mention this phenomenon, but provide few examples of different ways in which it might occur.[1] This is unfortunate, because expositing examples of how this problem can arise, and what to do about it, can be an eye-opener for students, as well as a great help to practitioners struggling with this problem. The purpose of this paper is to fill this void in our textbook literature by gathering together several possible reasons for obtaining the "wrong" sign, and suggesting how corrections might be undertaken.

A wrong sign can be considered a blessing, not a disaster. Getting a wrong sign is a friendly message that some detective work needs to be done – there is undoubtedly some shortcoming in the researcher's theory, data, specification, or estimation procedure. If the "correct" signs had been obtained, odds are that the analysis would not be double-checked. The following examples provide a checklist for this double-checking task, many illustrating substantive improvements in specification.

1. **Bad Economic Theory**. Suppose you are regressing the demand for Ceylonese tea on income, the price of Ceylonese tea and the price of Brazilian coffee. To your surprise you get a positive sign on the price of Ceylonese tea. This dilemma is resolved by recognizing that it is the price of other tea, such as Indian tea, that is the relevant substitute here. Rao and Miller (1971, p.38-9) provide this example. Gylfason (1981) refers to many studies which obtained "wrong" signs because they used the nominal rather than real interest rate when explaining consumption spending.

[1] Wooldridge (2000) is an exception; several examples of wrong signs are scattered throughout this text.

2. **Omitted Variable**. Suppose you are running an hedonic regression of automobile prices on a variety of auto characteristics such as horsepower, automatic transmission, and fuel economy, but keep discovering that the estimated sign on fuel economy is negative. Ceteris paribus, people should be willing to pay more, not less, for a car that has higher fuel economy, so this is a "wrong" sign. An omitted explanatory variable may be the culprit. In this case, we should look for an omitted characteristic that is likely to have a positive coefficient in the hedonic regression, but which is negatively correlated with fuel economy. Curbweight is a possibility, for example. (Alternatively, we could look for an omitted characteristic which has a negative coefficient in the hedonic regression and is positively correlated with fuel economy.) Here is another example, in the context of a probit regression. Suppose you are using a sample of females who have been asked whether they smoke, and then are resampled twenty years later. You run a probit on whether they are still alive after twenty years, using the smoking dummy as the explanatory variable, and find to your surprise that the smokers are more likely to be alive! This could happen if the non-smokers in the sample were mostly older, and the smokers mostly younger, reflecting Simpson's paradox. Adding age as an explanatory variable solves this problem, as noted by Appleton, French, and Vanderpump (1996).
3. **High Variances**. Suppose you are estimating a demand curve by regressing quantity of coffee on the price of coffee and the price of tea, using time series data, and to your surprise find that the estimated coefficient on the price of coffee is positive. This could happen because over time the prices of coffee and tea are highly collinear, resulting in estimated coefficients with high variances – their sampling distributions will be widely spread, and may straddle zero, implying that it is quite possible that a draw from this distribution will produce a "wrong" sign. Indeed, one of the casual indicators of multicollinearity is the presence of "wrong" signs. In this example, a reasonable solution to this problem is to introduce additional information by using the ratio of the two prices as the explanatory variable, rather than their levels. This example is one in which the wrong sign problem is solved by incorporating additional information to reduce

high variances. Multicollinearity is not the only source of high variances, however; they could result from a small sample size, or minimal variation in the explanatory variables. Leamer (1978, p.8) presents another example of how additional information can solve a wrong sign problem. Suppose you regress household demand for oranges on total expenditure E, the price p_o of oranges, and the price p_g of grapefruit (all variables logged), and are surprised to find wrong signs on the two price variables. Impose homogeneity, so that if prices and expenditure double, the quantity of oranges purchased should not change; this implies that the sum of the coefficients of E, p_o, and p_g is zero. This extra information reverses the price signs.

4. **Selection Bias**. Suppose you are regressing academic performance, as measured by SAT scores (the scholastic aptitude test is taken by many students to enhance their chances of admission to the college of their choice) on per student expenditures on education, using aggregate data on states, and discover that the more money the government spends, the less students learn! This "wrong" sign may be due to the fact that the observations included in the data were not obtained randomly – not all students took the SAT. In states with high education expenditures, a larger fraction of students may take the test. A consequence of this is that the overall ability of the students taking the test may not be as high as in states with lower education expenditure and a lower fraction of students taking the test. Some kind of correction for this selection bias is necessary. In this example, putting in the fraction of students taking the test as an extra explanatory variable should work. This example is taken from Guber (1999). Currie and Cole (1993) exposit another good example of selection bias. Suppose you are regressing the birthweight of children on several family and background characteristics, including a dummy for participation in AFDC (aid for families with dependent children), hoping to show that the AFDC program is successful in reducing low birthweights. To your consternation the slope estimate on the AFDC dummy is negative! This probably happened because mothers self-selected themselves into this program – mothers believing they were at risk for delivering a low birthweight child may have been more likely to participate in AFDC. This could

be dealt with by using the Heckman two-stage correction for selection bias or an appropriate maximum likelihood procedure. A possible alternative solution is to confine the sample to mothers with two children, for only one of which the mother participated in the AFDC program. A panel data method such as fixed effects (or differences) could be used to control for the unobservables that are causing the problem.

5. **Data Definitions/Measurement Error**. Suppose you are regressing stock price changes on a dummy for bad weather, in the belief that bad weather depresses traders and they tend to sell, so you expect a negative sign. But you get a positive sign. Rethinking this, you change your definition of bad weather from 100 percent cloud cover plus relative humidity above 70 percent, to cloud cover more than 80% or relative humidity outside the range 25 to 75 percent. Magically, the estimated sign changes. This example illustrates more than the role of variable definitions/measurement in affecting coefficient signs – it illustrates the dangers of data mining and underlines the need for sensitivity analysis. This example appears in Kramer and Runde (1997). This is not the only way in which measurement problems can contribute to generating a wrong sign. It is not uncommon to regress the crime rate on the per capita number of police and obtain a positive coefficient, suggesting that more police engender more crime. One possible reason for this is that having extra police causes more crime to be reported. Another reason for how measurement error can cause a wrong sign is exposited by Bound, Brown, and Mathiowetz (2001). They document that often measurement errors are correlated with the true value of the variable being measured (contrary to the usual econometric assumption) and show how this can create extra bias sufficient to change a coefficient's sign.

6. **Outliers**. Suppose you are regressing infant mortality on doctors per thousand population, using data on the 50 US states plus the District of Columbia, but find that the sign on doctors is positive. This could happen because the District of Columbia is an outlier – relative to other observations, it has large numbers of doctors, and pockets of extreme poverty. Removing the outlier should solve the

sign dilemma. This example appears in Wooldridge (2000, p.303-4). Rowthorn (1975) points out that a nice OECD cross-section regression confirming Kaldor's law resulted from a random scatter of points and an outlier, Japan.

7. **Simultaneity/Lack of Identification**. Suppose you are regressing quantity of an agricultural product on price, hoping to get a positive coefficient because you are interpreting it as a supply curve. Historically, such regressions produced negative coefficients and were interpreted as demand curves – the exogenous variable "weather" affected supply but not demand, rendering this regression an identified demand curve. Estimating an unidentified equation would produce estimates of an arbitrary combination of the supply and demand equation coefficients, and so could be of arbitrary sign. The lesson here is check for identification. A classic example here is Moore (1914) who regressed quantity of pig iron on price, obtained a positive coefficient and announced a new economic discovery – an upward-sloping demand curve. He was quickly rebuked for confusing supply and demand curves. Morgan (1990, chapter 5) discusses historical confusion on this issue. The generic problem here is simultaneity. More policemen may serve to reduce crime, for example, but higher crime will cause municipalities to increase their police force, so when crime is regressed on police, it is possible to get a positive coefficient estimate. Identification is achieved by finding a suitable instrumental variable. This suggests yet another reason for a wrong sign – using a bad instrument.

8. **Bad Instruments**. Instrumental variable (IV) estimation is usually employed to alleviate the bias caused by correlation between an explanatory variable and the equation error. Suppose you are regressing incidence of violent crime on percentage of population owning guns, using data on U.S. cities. Because you believe that gun ownership is endogenous (i.e., higher crime causes people to obtain guns), you use gun magazine subscriptions as an instrumental variable for gun ownership and estimate using two-stage least squares. You have been careful to ensure identification, and check that the correlation between gun ownership and gun magazine subscriptions is substantive, so are very surprised to find that the IV

slope estimate is negative, the reverse of the sign obtained using ordinary least squares. This was caused by negative correlation between gun subscriptions and crime. The instrumental variable gun subscriptions was representing gun ownership which is culturally patterned, linked with a rural hunting subculture, and so did not represent gun ownership by individuals residing in urban areas, who own guns primarily for self-protection.[2] Another problem with IV estimation is that if the IV is only weakly correlated with the endogenous variable for which it is serving as an instrument, the IV estimate is not reliable and so a wrong sign could result.

9. **Specification Error**. Suppose you have student scores on a pretest and a posttest and are regressing their learning, measured as the difference in these scores, on the pretest score (as a measure of student ability), a treatment dummy (for some students having had an innovative teaching program) and other student characteristics. To your surprise the coefficient on pretest is negative, suggesting that better students learn less! Becker and Salemi (1977) spell out several ways in which specification bias could cause this. One example is that the true specification may be that the posttest score depends on the pretest score with a coefficient less than unity. Subtracting pretest from both sides of this relationship produces a negative coefficient on pretest in the relationship connecting the score difference to the pretest score. Measurement error could also be playing a role here. A positive measurement error in pretest appears negatively in the score difference, creating a negative correlation between the pretest explanatory variable and the equation error term, creating bias.

10. **Ceteris Paribus Confusion**. Suppose you have regressed house price on square feet, number of bathrooms, number of bedrooms, and a dummy for a family room, and are surprised to find the family room coefficient has a negative sign. The coefficient on the family room dummy tells us the change in the house price if a family room is added, holding constant the other regressor values, in particular holding constant square feet. So adding a family room under this constraint must

[2] I am indebted to Tomislav Kovandzic for this example.

entail a reduction in square footage elsewhere, such as smaller bedrooms or loss of a dining room, which will entail a loss in house value. In this case the net effect on price is negative. This problem is solved by asking what will happen to price if, for example, a 600 square foot family room is added, so that the proper calculation of the value of the family room involves a contribution from both the square feet regressor coefficient and the family room dummy coefficient. As another example, suppose you are regressing yearling (racehorse) auction prices on various characteristics of the yearling, plus information on its sire (father) and dam (mother). To your surprise you find that although the estimated coefficient on dam dollar winnings is positive, the coefficient on number of dam wins is negative, suggesting that yearlings from dams with more race wins are worth less. This wrong sign problem is resolved by recognizing that the sign is misinterpreted. In this case, the negative sign means that holding dam dollar winnings constant, a yearling is worth less if its dam required more wins to earn those dollars. Although proper interpretation solves the sign dilemma, in this case an adjustment to the specification seems appropriate: replace the two dam variables with a new variable, earnings per win. This example is taken from Robbins and Kennedy (2001).

11. **Interaction Terms**. Suppose you are regressing economics exam scores on grade point average (GPA) and an interaction term which is the product of GPA and ATTEND, percentage of classes attended. The interaction term is included to capture your belief that attendance benefits better students more than poorer students. Although the estimated coefficient on the interaction term is positive, as you expected, to your surprise the estimated coefficient on GPA is negative, suggesting that students with higher ability, as measured by GPA, have lower exam scores. This dilemma is easily explained – the partial derivative of exam scores with respect to GPA is the coefficient on GPA plus the coefficient on the interaction term times ATTEND. The second term probably outweighs the first for all ATTEND observations in the data, so the influence of GPA on exam scores is positive, as expected. Wooldridge (2000, p.190-1) presents this example.

12. **Regression to the Mean**. Suppose you are testing the convergence hypothesis by regressing average annual growth over the period 1950-1979 on GDP per work hour in 1950. Now suppose there is substantive measurement error in GDP. Large underestimates of GDP in 1950 will result in low GDP per work hour, and at the same time produce a higher annual growth rate over the subsequent period (because the 1979 GDP measure will likely not have a similar large underestimate). Large overestimates will have an opposite effect. As a consequence, your regression is likely to find convergence, even when none exists. This is a type of “wrong” sign, in this case produced by the regression to the mean phenomenon. For more on this example, see Friedman (1992). A similar example is identified by Hotelling (1933). Suppose you have selected a set of firms with high business-to-sales ratios and have regressed this measure against time, finding a negative relationship i.e., over time the average ratio declines. This result is likely due to the reversion to the mean phenomenon – the firms chosen probably had high ratios by chance, and in subsequent years reverted to a more normal ratio.

13. **Nonstationarity**. Regressing a random walk on an independent random walk should produce a slope coefficient insignificantly different from zero, but far too frequently does not, as is now well-known. This spurious correlation represents a “wrong” sign – the sign should not be significantly positive or negative. This is a very old problem, identified by Yule (1926) in an article entitled “Why do we sometimes get nonsense correlations between time series?”

14. **Common Trends.** A common trend could swamp what would otherwise be a negative relationship between two variables; omitting the common trend would give rise to the wrong sign.

15. **Functional Form Approximation**. Suppose you are running an hedonic regression of house prices on several characteristics of houses, including number of rooms and the square of the number of rooms. Although you get a positive coefficient on the square of number of rooms, to your surprise you get a negative

coefficient on number of rooms, suggesting that for a small number of rooms more rooms decreases price. This could happen because in your data there are no (or few) observations with a small number of rooms, so the quadratic term dominates the linear term throughout the range of the data. The negative sign on the linear term comes about because it provides the best approximation to the data. Wooldridge (2000, p.188) provides this example.

16. **Dynamic Confusion**. Suppose you have regressed income on lagged income and investment spending. You are interpreting the coefficient on investment as the multiplier and are surprised to find that it is less than unity, a type of "wrong sign." Calculating the long-run impact on income this implies, however, resolves this dilemma. This example appears in Rao and Miller (1971, p.44-5). Suppose you have panel data on the US states and are estimating the impact of public capital stock (in addition to private capital stock and labor input) on state output. You estimate using fixed effects and to your surprise obtain a negative sign on the public capital stock coefficient estimate. Baltagi and Pinnoi (1995) note that this could be because fixed effects estimates the short-run reaction; pooled OLS, the "between" estimator, and random effects all produce the expected positive sign, suggesting that the long-run impact is positive. Suppose you believe that x affects y positively but there is a lag involved. You regress y_t on x_t and x_{t-1} and are surprised to find a negative coefficient on x_{t-1}. The explanation for this is that the long-run impact of x is smaller than its short-run impact.

17. **Reversed Measure**. Suppose you are regressing consumption on a consumer confidence measure, among other variables, and unexpectedly obtain a negative sign. This could happen because you didn't realize that small numbers for the consumer confidence measure correspond to high consumer confidence. It has been known[3] for an economist to present an entire seminar trying to explain a wrong sign only to discover afterwards that it resulted from his software reversing the coding on his logit analysis.

---

[3] I am indebted to Marie Rekkas for this anecdote.

18. **Heteroskedasticity**. Suppose you are estimating a probit model, with the latent equation a linear function of x, namely $y^* = \alpha + \beta x + \varepsilon$, but the error ε is heteroskedastic, with variance σ^2 proportional to the square of x. Probit estimates β/σ, not β, because the likelihood function is based on the cumulative *standard* normal density. So the operative latent equation is proportional to $\alpha/x + \beta$, in which the influence of x is reversed in sign. See Wooldridge (2001, p.479) for discussion.

19. **Underestimated Variances.** If the variance of a coefficient estimate is underestimated, an irrelevant variable could be statistically "significant," of either sign. The Poisson model assumes that the variance of the counts is equal to its expected value. Because of this Poisson estimation produces marked underestimates of coefficient estimates' variances in the typical case in which there is overdispersion (the count variance is larger than its expected value). Researchers often rely on asymptotic properties of test statistics which could be misleading in small samples. A classic example appears in Laitinen (1978) who showed that failure to use small-sample adjustments explained why demand homogeneity had been rejected so frequently in the literature.

What should be done if your double-checking can turn up no reasonable explanation for the "wrong" sign? Try and get it published. Wrong sign puzzles, such as the Leontief paradox, are a major stimulus to the development of our discipline. For example, recent evidence suggests that there is a positive relationship between import tariffs and growth across countries in the late 19th century, a "wrong" sign in many economists' view. Irwin (2002) extends the relevant economic theory to offer an explanation for this.

There is no definitive list of ways in which "wrong" signs can be generated. In general, any theoretical oversight, specification error, data problem, or inappropriate estimating technique could give rise to a "wrong" sign. Observant readers might have noted that many could be classified under a single heading: Researcher Foolishness. This serves to underline the importance of the first of Kennedy's (2002) ten commandments of applied econometrics: Use Common Sense.

REFERENCES

Appleton, D. R., J. M. French, and M. P. J. Vanderpump. 1996. Ignoring a Covariate: An Example of Simpson's Paradox. *American Statistician* 50: 340-1.

Baltagi, B. H. and N. Pinnoi. 1995. Public Capital Stock and State Productivity Growth: Further Evidence from an Error Components Model. *Empirical Economics* 20: 351-9.

Becker, W. E. and M. K. Salemi. 1977. The Learning and Cost Effectiveness of AVT Supplemented Instruction: Specification of Learning Models. *Journal of Economic Education* 8: 77-92.

Bound, J., C. Brown and N. Mathiowetz. 2001. Measurement Error in Survey Data. In J. J. Heckman and E. E. Leamer (eds), *Handbook of Econometrics*, vol.V. Amsterdam: North Holland, 3705-3843.

Currie, J. and N. Cole. 1993. Welfare and Child Health: The Link between AFDC Participation and Birth Weight. *American Economic Review* 83: 971-83.

Friedman, M. 1992. Do Old Fallacies Ever Die? *Journal of Economic Literature* 30: 2129-32.

Guber, D. C. 1999. Getting What You Pay For: The Debate over Equity in Public School Expenditures. *Journal of Statistics Education* 7(2). [Online] Available at: www.amstat.org/publications/jse/secure/v7n2/datasets.guber.cfm

Gylfason, H. 1981. Interest Rates, Inflation, and the Aggregate Consumption Function. *Review of Economics and Statistics* 63: 233-45.

Hotelling, H. 1933. Review of *The Triumph of Mediocrity in Business,* by Horace Secrist. *Journal of the American Statistical Association* 28: 463-5.

Irwin, D. A. 2002. Interpreting the Tariff-Growth Correlation of the Late 19th Century. *American Economic Review, Papers and Proceedings* 92: 165-9.

Kennedy, P. E. 2002. Sinning in the Basement: What are the Rules? The Ten Commandments of Applied Econometrics. *Journal of Economic Surveys*, 16:

Kramer, W. and R. Runde. 1997. Stocks and the Weather: An Exercise in Data Mining or Yet Another Capital Market Anomaly? *Empirical Economics* 22: 637-41.

Laitinen, K. (1978) Why is Demand Homogeneity So Often Rejected? *Economics Letters* 1: 187-91.

Leamer, E. E. 1978. *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. New York: John Wiley.

Morgan, M. S. (1990) *The History of Econometric Ideas*. Cambridge: Cambridge University Press.

Moore, H. L. (1914) *Economic Cycles – Their Law and Cause*. New York: Macmillan.

Rao, P. and R. Miller (1971) *Applied Econometrics*. Belmont, CA: Wadsworth.

Robbins, M. and P. E. Kennedy. 2001. Buyer Behaviour in a Regional Thoroughbred Yearling Market, *Applied Economics* 33: 969-77.

Rowthorn, R. E. 1975. What remains of Kaldor's Law? *Economic Journal* 85: 10-19.

Smith, G. and W. Brainard. 1976. The Value of A Priori Information in Estimating a Financial Model. *Journal of Finance* 31: 1299–322.

Wooldridge, J. M. (2000) *Introductory Econometrics*. Cincinnati: South-Western.

Wooldridge, J. M. (2002) *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Mass.: MIT Press.

Yule, G. U. 1926. Why Do We Sometimes Get Nonsense Correlations Between Time-Series? *Journal of the Royal Statistical Society* 89: 1-64.