to the following formula for the kernel density estimator:

$$f(x) = (1/hN)\Sigma K[(x_i - x)/h],$$

where K is the kernel function.

- In the specification $y = m(x) + \varepsilon$, $m(x)$ is the conditional mean of y, namely the mean of the distribution of y given x. From elementary statistics this distribution, written $f(y|x)$, is equal to $f(x,y)/f(x)$ where $f(x,y)$ is the joint distribution of x and y, and $f(x)$ is the marginal distribution of x. A kernel smoothing method can be used to estimate these two distributions and so $f(y|x)$ can be estimated as their ratio. Estimating the mean of this distribution creates an estimate of $m(x)$. This estimate turns out to be identical to that given earlier in the general notes, namely

$$\hat{m}(x) = \Sigma y_i K[(x_i - x)/h]/\Sigma K[(x_i - x)/h],$$

offering formal justification for the use of this weighted averaging procedure.

- The variance of the error term in the specification $y = m(x) + \varepsilon$ is the variance of y conditional on x. For a homoskedastic error this would be the same for all values of x. In general terms, this variance is given by

$$\begin{aligned} V(y \mid x) &= \sigma^2(\varepsilon) = E[(y - E(y))^2 \mid x] \\ &= E(y^2 \mid x) - [E(y \mid x)]^2. \end{aligned}$$

Two ways of estimating this are evident.

First, $[E(y|x)]^2$ can be estimated by the square of $\hat{m}(x)$ given earlier, and $E(y^2|x)$ can be estimated by using the formula for $\hat{m}(x)$ with y_i replaced by y_i^2.

Second, we can estimate ε_i by $y_i - \hat{m}(x_i)$ and then use the formula for $\hat{m}(x)$ with y_i replaced by the square of this estimated ε_i.

Chapter 22

Applied Econometrics

22.1 Introduction

The preceding chapters have focused on econometric theory. Unfortunately, unpleasant realities of real-world data force applied econometricians to violate the prescriptions of econometric theory as taught by our textbooks. Leamer (1978, p. vi) vividly describes this behavior as wanton sinning:

> As it happens, the econometric modeling was done in the basement of the building and the econometric theory courses were taught on the top floor (the third). I was perplexed by the fact that the same language was used in both places. Even more amazing was the transmogrification of particular individuals who wantonly sinned in the basement and metamorphosed into the highest of high priests as they ascended to the third floor.

It is no secret that there is a world of difference between applied and theoretical econometrics. In fact, there is a remarkable lack of communication between econometric theorists and applied econometricians – the former, who are often called upon to teach applied econometrics courses, are notorious for teaching econometric theory in these courses. (Examples are given, and an applied paper is usually required, to justify calling the course an applied econometrics course!)

In these "applied" courses students typically are taught, in hands-on fashion, how to undertake a wide variety of econometric techniques. Examples at the elementary level are the use and interpretation of dummy variables, the logic of F and chi-square tests, and testing and correcting for nonspherical errors. Examples at a more advanced level are testing for unit roots and cointegration, correcting for sample selection bias, performing Hausman tests, and estimating using Tobit, Poisson, and ordered probit models. But the focus is on the mechanics of estimation and testing rather than on the fundamentals of applied work such as problem articulation, data cleaning, and model specification. In short, teaching is technique oriented rather than problem oriented.

Why is this? There are several reasons. First, teaching applied econometrics is dif ficult, because doing applied work is difficult, or, in blunter terms, econometrics i much easier without data. Second, because doing quality applied work brings littl prestige in the profession, many econometrics instructors have never done any applie work. Third, in common with other specialists, econometricians teach what they enjc teaching and what they know how to teach, not what students need.

What do students "need" in the context of applied econometrics? Because tl "sinning" identified by Leamer is inevitable, students need to learn some standa operating procedures, or rules of behavior, which if applied will bound this sinnir as well as help avoid elementary mistakes. Most econometrics instructors believe th students do not need to be taught such elementary rules of behavior, and in any eve they would be uncomfortable teaching them because these rules do not have the int lectual rigor and clean answers so prized by econometric theorists. This is one reas why such rules seldom appear in textbooks or course lectures.

The teaching task is unquestionably difficult. Tukey (1969, p. 90) expresses t difficulty eloquently:

> Divert them from depending upon the "authority" of standard textbook solutions, without being able to substitute a second religion for the first. Stimulate intelligent prob formulation, without being able to say quite how this is done. Demand high standard statistical reasoning, but without specifying a simple model of statistics which might serv a criterion for quality of reasoning.

In keeping with the flavor of this book, the purpose of this chapter is to discu variety of practical dimensions of doing applied econometric work that are often n ing from applied econometrics courses, such as rules of behavior for bounding sin and avoiding elementary mistakes. As the Tukey quote suggests, however, readers ing to find a definitive methodology will be disappointed.

22.2 The Ten Commandments of Applied Econometrics

One of the most prevalent sins of applied econometricians is mechanical applic of rules and procedures in inappropriate settings. The rules of behavior sugg below have a different flavor, however; they are designed to force practition avoid mechanical application of econometric techniques learned in econometric t courses, and so escape this sin.

Rule 1: Use common sense and economic theory

The reason for this rule is that common sense is not all that common. Indeed, som it appears that not much thought has gone into empirical work, let alone good th Nor does such thought require complicated econometrics. For example, this th should cause researchers to match per capita variables with per capita variabl real exchange rates to explain real imports/exports, employ nominal interest r explain real money demand, select appropriate functional forms for dependent va constrained to lie between zero and one, resist trying to explain a trendless variable with a trended variable, avoid cardinalizing ordered qualitative explanatory variables, beware of the regression fallacy, and never infer causation from correlation.

Rule 2: Avoid type III errors

A type III error occurs when a researcher produces the right answer to the wrong question. A corollary of this rule is that an approximate answer to the right question is worth a great deal more than a precise answer to the wrong question.

The phenomenon at issue here is that the relevant objective/hypothesis/specification may be completely different from what is initially suggested. Econometricians experience this regularly when colleagues or students stop by for advice, prefacing their request with words to the effect that they do not want to take up much of the econometrician's time so they will explain just the technical detail with which they want help. Acquiescing to this is usually a mistake because more often than not asking simple questions about the context of the problem brings to light serious misunderstandings. For example, it may be that it is the cumulative change in a variable that is relevant, not the most recent change, or that the hypothesis under test should be that a coefficient is equal to another coefficient, rather than equal to zero, or that the dependent variable observations are durations, so that a duration model should be used.

The main lesson here is a blunt one: Ask questions, especially seemingly foolish questions, to ensure that you have a full understanding of the context of the "technical detail" being discussed; often it turns out that the research question has not been formulated appropriately.

Rule 3: Know the context

This rule is a natural extension of the previous rule. It is crucial that one becomes intimately familiar with the phenomenon being investigated – its history, institutions, operating constraints, measurement peculiarities, cultural customs, and so on, going beyond a thorough literature review. Again, questions must be asked: Exactly how were the data gathered? Did government agencies impute the data using unknown formulas? What were the rules governing the auction? How were the interviewees selected? What instructions were given to the participants? What accounting conventions were followed? How were the variables defined? What is the precise wording of the questionnaire? How closely do measured variables match their theoretical counterparts? Another way of viewing this rule is to recognize that you, the researcher, know more than the computer – you know, for example, that water freezes at zero degrees Celsius, that people tend to round their incomes to the nearest five thousand, and that some weekends are 3-day weekends.

Rule 4: Inspect the data

Even if a researcher knows the context, he or she needs to become intimately familiar with the specific data with which he or she is working. Economists are particularly prone to the complaint that researchers do not know their data very well, a phenomenon made worse by the computer revolution, allowing researchers to obtain and work with data electronically by pushing buttons.

Inspecting the data involves summary statistics, graphs, and data cleaning, to both check and "get a feel for" the data. Summary statistics can be very simple, such as calculating means, standard errors, maximums, minimums, and correlation matrices, or more complicated, such as computing condition indices and influential observation diagnostics.

The advantage of graphing is that a picture can force us to notice what we never expected to see. Researchers should supplement their summary statistics with simple graphs: histograms, residual plots, scatterplots of residualized data, and graphs against time. It is important to graph the data in several different ways.

Data cleaning looks for inconsistencies in the data – are any observations impossible, unrealistic, or suspicious? The questions here are mostly simple, but could become more complicated in a particular context. Do you know how missing data were coded? Are dummies all coded zero or one? Are all observations consistent with applicable minimum or maximum values? Do all observations obey logical constraints they must satisfy?

The main message of this rule is that instead of beginning by asking "what technique should I use here?," practitioners should ask "how can I summarize and understand the main features of these data?"

Rule 5: Keep it sensibly simple

This KISS rule should not be confused with the commercial "keep it simple, stupid" rule, because some simple models are stupid, containing logical errors or being at variance with facts. As in science, progress in economics results from beginning with simple models, seeing how they work in applications, and then modifying them if necessary. Examples are the functional form specifications of some Nobel Laureates – Tinbergen's social welfare functions, Arrow's and Solow's work on the CES production function, Friedman's, Becker's, Tobin's, and Modigliani's consumer models, and Lucas's rational expectations model.

Beginning with a simple model is referred to as following the bottom-up, or specific-to-general, approach to developing an econometric specification. Its main drawback is that testing is biased if the simple model does not nest the real-world process generating the data. But no such true model can ever be found, so this disadvantage is shared by the competing top-down, or general-to-specific, approach, albeit to a different degree. The main problem with the top-down approach is that it is not realistic to think that we can begin by estimating a general model incorporating all conceivable explanatory variables and functional forms. Because of this, applications of this method require that the researcher be able to think of the "right" general model from the start.

The top-down approach nonetheless has an attractive feature – testing is likely to be less biased. In light of this, a compromise methodology has evolved. Practitioners begin with simple models which are expanded whenever they fail. Failures are identified through misspecification tests such as evaluation of out-of-sample forecasts. Expansions are on the one hand modest in that they introduce one extra layer of complexity (a new variable, for example), but on the other hand quite general in that they cover a range of possible roles for the new element (generous lags, for example), as degrees of freedom allow. Testing down is undertaken to create a new simple model which is subjected to misspecification tests, and this process of discovery is repeated. In this way simplicity is combined with the general-to-specific methodology, producing a compromise process which, judged by its wide application, is viewed as an acceptable rule of behavior.

Rule 6: Use the interocular trauma test

Output from modern empirical work typically fills many pages, as researchers try a variety of functional forms and sets of explanatory variables. This rule cautions researchers to look long and hard at this plethora of results: Look at the results until the answer hits you between the eyes! Part of this rule is to check that the results make sense. Are the signs of coefficients as expected? Are important variables statistically significant? Are coefficient magnitudes reasonable? Are the implications of the results consistent with theory? Are there any anomalies? Are any obvious restrictions evident? Apply the "laugh" test – if the findings were explained to a layperson, could that person avoid laughing?

But another part of this rule is more subtle, and subjective. By looking long and hard at reams of computer output, a researcher should eventually, through both conscious and subconscious means, recognize the message they are conveying (which could be a negative message) and become comfortable with it. This subjective procedure should be viewed as separate from and complementary to formal statistical testing procedures used to investigate what is going on. Indeed, the results of such testing procedures form part of the mass of statistical output one is trying to interpret.

Rule 7: Understand the costs and benefits of data mining

As discussed in chapter 5, there are two variants of "data mining," one classified as the greatest of the basement sins, but the other viewed as an important ingredient in data analysis. Unfortunately, these two variants usually are not mutually exclusive and so frequently conflict in the sense that to gain the benefits of the latter, one runs the risk of incurring the costs of the former.

The undesirable version of data mining occurs when one tailors one's specification to the data, resulting in a specification that is misleading because it embodies the peculiarities of the particular data at hand. Furthermore, traditional testing procedures used to "sanctify" the specification are no longer legitimate, because these data, since they have been used to generate the specification, cannot be judged impartial if used to test that specification. The alternative version of "data mining" refers to experimenting with the data to discover empirical regularities that can inform economic theory. Its greatest virtue is that it can uncover empirical regularities that point to errors/omissions in theoretical specifications.

The process by which a specification is developed, blending economic theory, common sense, and a judicious mixture of both bottom-up and top-down, clearly incorporates elements of "data mining," a terminology with strong emotive content. Data mining is inevitable; the art of the applied econometrician is to allow for data-driven theory while avoiding the considerable dangers inherent in data mining.

Rule 8: Be prepared to compromise

In virtually every econometric analysis there is a gap, usually a vast gulf, between the problem at hand and the closest scenario to which standard econometric theory is applicable. Very seldom does one's problem even come close to satisfying the assumptions under which econometric theory delivers an optimal solution. A consequence of this is that practitioners are always forced to compromise and adopt suboptimal solutions, the characteristics of which are unknown. Leamer (1997, p. 552) lends this special emphasis when listing his choices for the three most important aspects of real data analyses: "compromise, compromise, compromise."

The issue here is that in their econometric theory courses students are taught standard solutions to standard problems, but in practice there are no standard problems, only standard solutions. Applied econometricians are continually faced with awkward compromises, and must be willing to make *ad hoc* modifications to standard solutions. Should a proxy be used? Can sample attrition be ignored? Should these unit root tests be believed? Is aggregation legitimate here?

Rule 9: Do not confuse statistical significance with meaningful magnitude

Very large sample sizes, such as those that have become common in cross-sectional data, thanks to the computer revolution, can give rise to estimated coefficients with very small standard errors. A consequence of this is that coefficients of trivial magnitude may test significantly different from zero, creating a misleading impression of what is important. Because of this, researchers must always look at the magnitude of coefficient estimates as well as their significance.

An even more serious problem associated with significance testing is that there is a tendency to conclude that finding significant coefficients "sanctifies" a theory, with a resulting tendency for researchers to stop looking for further insights. Sanctification via significance testing should be replaced by continual searches for additional evidence, both corroborating evidence and, especially, disconfirming evidence. If your theory is correct, are there testable implications? Can you explain a range of interconnected findings? Can you find a bundle of evidence consistent with your hypothesis but inconsistent with alternative hypotheses? Can your theory "encompass" its rivals in the sense that it can explain other models' results?

Rule 10: Report a sensitivity analysis

Econometricians base their analyses on an imaginary assumed "data-generating process" (DGP) which is viewed as having produced the data being used for estimation. More likely than not, this fictitious DGP does not correspond even closely to the way in which the data actually were generated. Because of this, it is important to check if the empirical results are sensitive to the assumptions upon which the estimation has been based. This is the purpose of a "sensitivity analysis," indicating to what extent the substantive results of the research are affected by adopting different specifications about which reasonable people might disagree. For example, are the results sensitive to the sample period, the functional form, the set of explanatory variables, or measurement of or proxies for the variables? Are robust estimation results markedly different? The context of the problem should indicate what issues are most important check, but in general such decisions are never easy. In this respect the problem is similar to that described earlier in rule #8 when compromising.

There exists a second dimension to sensitivity analyses. Published research papers are typically notoriously misleading accounts of how the research was actually conducted. Because of this it is very difficult for readers of research papers to judge the extent to which data mining may have unduly influenced the results. Indeed, results tainted by subjective specification decisions undertaken during the heat of econometric battle should be considered the rule rather than the exception. When reporting a sensitivity analysis, researchers should explain fully their specification search so that readers can judge for themselves how the results may have been affected. This is basically an "honesty is the best policy" approach, advocated by Leamer (1978, p. vi): "Sinners are not expected to avoid sins; they need only confess their errors openly."

These rules have given rise to the ten commandments of applied econometrics.

1. **Thou shalt use common sense and economic theory.**
   Corollary: Thou shalt not do thy econometrics as thou sayest thy prayers.
2. **Thou shalt ask the right questions.**
   Corollary: Thou shalt place relevance before mathematical elegance.
3. **Thou shalt know the context.**
   Corollary: Thou shalt not perform ignorant statistical analyses.
4. **Thou shalt inspect the data.**
   Corollary: Thou shalt place data cleanliness ahead of econometric godliness.
5. **Thou shalt not worship complexity.**
   Corollary: Thou shalt not apply asymptotic approximations in vain.
   Corollary: Thou shalt not talk Greek without knowing the English translation.
6. **Thou shalt look long and hard at thy results.**
   Corollary: Thou shalt apply the laugh test.
7. **Thou shalt beware the costs of data mining.**
   Corollary: Thou shalt not worship R^2.
   Corollary: Thou shalt not hunt statistical significance with a shotgun.
   Corollary: Thou shalt not worship the 5% significance level.
8. **Thou shalt be willing to compromise.**
   Corollary: Thou shalt not worship textbook prescriptions.
9. **Thou shalt not confuse significance with substance.**
   Corollary: Thou shalt not ignore power.
   Corollary: Thou shalt not test sharp hypotheses.
   Corollary: Thou shalt seek additional evidence.
10. **Thou shalt confess in the presence of sensitivity.**
    Corollary: Thou shalt anticipate criticism.

Knowing these ten commandments is not enough to guarantee quality applied work – inspecting the data requires knowing how to inspect, what to look for, and how to interpret what is found, not to mention remembering to look; the interocular trauma test seems trivial, but is hard to perform; knowing that it is necessary to compromise does not mean that a researcher knows how to compromise. Much of the skill of the applied econometrician is judgmental and subjective, characterized in the literature as "lore" or "tacit knowledge." This "lore" can only be learned by experience and by watching the masters, too much to expect this chapter or a course in applied econometrics to accomplish.

Accordingly, the remainder of this chapter will not try to instill in readers this lore, but instead will attempt the much more modest task of alerting readers to common technical errors made by practitioners, surveying a broad range of techniques and applications of which all researchers should be aware, and discussing what for many is a major trauma – getting the wrong sign.

22.3 Getting the Wrong Sign

A remarkably common occurrence when doing applied work is to run an *a priori* favor ite specification and discover a "wrong" sign. Rather than considering this a disaster a researcher should consider it a blessing – this result is a friendly message that som detective work needs to be done – there is undoubtedly some shortcoming in one's the ory, data, specification, or estimation procedure. If the "correct" signs had been obtained odds are that the analysis would not be double-checked. What should be checked?

The first step is always to check economic theory. It is amazing how after the fac economists can conjure up rationales for incorrect signs. But one should never sto here. If there was good reason *a priori* to expect a different sign, there is a moral obli gation to check econometric reasons for why the "wrong" sign was obtained, befor changing the theory. Here is a top ten list of econometric reasons for "wrong" signs.

1. *Omitted variable* Suppose you are running a hedonic regression of automobil prices on a variety of auto characteristics such as horsepower, automatic tran mission, and fuel economy, but keep discovering that the estimated sign on fu economy is negative. *Ceteris paribus*, people should be willing to pay more, n less, for a car that has higher fuel economy, so this is a "wrong" sign. An omitt explanatory variable may be the culprit. In this case, we should look for an omitt characteristic that is likely to have a positive coefficient in the hedonic regressio but which is negatively correlated with fuel economy. Curbweight is a possibili for example. Alternatively, we could look for an omitted characteristic which h a negative coefficient in the hedonic regression and is positively correlated w fuel economy.

   Suppose you are using a sample of females who have been asked whether th smoke, and then are resampled 20 years later. You run a probit on whether th are still alive after 20 years, using the smoking dummy as the explanatory va able, and find to your surprise that the smokers are more likely to be alive! Th could happen if the nonsmokers in the sample were mostly older, and the smok mostly younger. Adding age as an explanatory variable solves this problem.
2. *High variances* Suppose you are estimating a demand curve by regressing qua tity of coffee on the price of coffee and the price of tea, using time series data, a to your surprise find that the estimated coefficient on the price of coffee is posi

This could happen because 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 use the ratio of the two prices as the explanatory variable rather than their levels.

Multicollinearity is not the only source of high variances; they could result from a small sample size, or minimal variation in the explanatory variables. Suppose you regress household demand for oranges on total expenditure, the price of oranges, and the price of grapefruit (all variables logged), and are surprised to find wrong signs on the two price variables. Impose homogeneity, namely that if prices and expenditure double, the quantity of oranges purchased should not change; this implies that the sum of the coefficients of expenditure and the two price variables is zero. Incorporation of this extra information could reverse the price signs.

3. *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. When using individual data, the Heckman two-stage correction for selection bias or an appropriate maximum likelihood procedure would be in order.

   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 of delivering a low birthweight child may were more likely to participate in AFDC.
4. *Ceteris paribus confusion* Suppose you are regressing yearling (year-old race-horse) auction prices on various characteristics of the yearling, plus information on their sires (fathers) and dams (mothers). 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.

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 must 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.

5. *Data definitions/measurement* 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% cloud cover plus relative humidity above 70%, to cloud cover more than 80% or relative humidity outside the range 25–75%. 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.

   A common example of the influence of measurement problems occurs when a regression of the crime rate on the per capita number of police turns up a positive coefficient, suggesting that more police engender more crime! What may be happening here is that having extra police causes more crime to be reported. An alternative explanation, noted later, is that more crime may be causing the authorities to hire more police.

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. If, as is the case here, the outlying observation is such that it is not representative, it should be removed.

7. *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 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.

8. *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. A specification error could have caused this. For example, 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.
9. *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 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. As discussed in chapter 11, identification is achieved by finding a suitable instrumental variable. This suggests yet another reason for a wrong sign – using a bad instrument.
10. *Bad instrument* Suppose you are regressing incidence of violent crime on percentage of population owning guns, using data on US 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 problem was solved when it was discovered that the correlation between gun subscriptions and crime was negative. 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.

As these examples (and others in the general and the technical notes to this section) illustrate, the value of finding a "wrong" sign is that it can prompt development of a better specification. In some cases this reflects the good variant of data mining (identifying an outlier, discovering an omitted variable, awakening to selection bias, using relative prices, or adopting earnings per win), but in other cases it illustrates the bad variant of data mining (changing the definition of bad weather to suit one's needs). In all cases they require application of the ten commandments.

22.4 Common Mistakes

Failure to apply the ten commandments is the biggest source of mistakes made by practitioners – failing to use economic theory, abandoning common sense, addressing the wrong question, not knowing the context, never checking the data, making things too complicated, not looking/thinking long and hard about the results, molding the specification to the data, not learning from the data, paying undue attention to significance tests, and forgetting to perform a sensitivity analysis. In addition to these fundamental errors, however, there are several technical mistakes frequently made by practitioners that can easily be avoided. Here is a top baker's-dozen list.

1. Interpretation of a significant DW test or heteroskedasticity test as pointing to the need to change estimation technique from OLS to EGLS. It should initially be interpreted as "something is wrong with the specification."
2. Thinking that White's heteroskedasticity-consistent estimation produces coefficient estimates different from those produced by OLS. In this procedure coefficient estimates are unchanged from OLS; it is the estimate of the variance–covariance matrix that is different.
3. Forgetting interaction or quadratic terms when assessing variable influence. The role of an explanatory variable in affecting the dependent variable is given by the derivative of the dependent variable with respect to that explanatory variable, which may not be just the coefficient on that variable.
4. Using a linear functional form when the dependent variable is a fraction. A linear functional form could be an adequate approximation if no observations are near zero or unity; otherwise a logistic functional form, in which the log odds ratio is regressed on the explanatory variables, is probably more suitable.
5. Believing that multicollinearity creates bias, or invalidates inference. No bias is created by multicollinearity. Although estimated variances are large, they are unbiased estimates of a large variance, and so inference is unaffected – the type I error is what it has been selected to be.
6. Solving multicollinearity between variables X and W by residualizing W for X (i.e., removing the linear influence of X on W) to get Wrx and then regressing Y on X and Wrx instead of regressing Y on X and W. As the Ballentine makes clear, this produces a biased estimate of the coefficient on X.
7. Using an ordered qualitative variable as a regressor. Consider a variable "education" coded one for some elementary school, two for some high school, three for some university, and so on. Using this variable as a regressor forces the impact on the dependent variable of moving from elementary school to high school to be the same as the impact of moving from high school to university, and so forth. Only if these implicit restrictions are tested and accepted should this variable be used as a regressor. Otherwise separate dummy variables for each category should be employed.
8. Measuring forecast success in logit/probit models by the fraction of outcomes predicted correctly. How does this compare to forecasting every observation as a one or forecasting every observation as a zero? Success is better measured by averaging the fraction of ones correctly forecast and the fraction of zeros correctly forecast.

9. Interpreting the LM test for a nonzero variance of "random" intercepts in panel data as a test for random effects versus fixed effects. This is a test for testing whether the intercepts are all equal, an alternative to the F test for testing equality of the fixed effects intercepts. A Hausman test is needed to test for the appropriateness of the random effects specification.
10. Using Tobit in a context in which it is clear that a separate equation should be used to determine limit observations. A classic example is when any expenditure involves considerable expenditure, so that zero is not the right limit to use in estimation.
11. Testing for unit roots without a strategy for determining if a drift or time trend should be included. The power of unit root tests can be enhanced by using subjective judgment concerning the need for a drift or time trend.
12. Not understanding selection bias, particularly self-selection bias. *Unobserved* characteristics of an individual may affect both the dependent variable *and* decisions made by that individual determining whether he or she is observed or to what dummy variable category he or she belongs. Adding the inverse Mills ratio works for linear regressions; it does not make sense for nonlinear regressions such as when estimating logit, probit, and count data models.
13. Forgetting about possible endogeneity in the empirical specification. Too often researchers do not think through this issue, resulting in empirical specifications with endogenous regressors. A consequence is that OLS is used when an alternative, such as IV estimation, may be more appropriate.

22.5 What do Practitioners Need to Know?

Practitioners need to follow the ten commandments, know how to deal with "wrong" signs, and avoid common mistakes. But they also need to know a wide range of econometric techniques and when to apply them. What are the basic things practitioners need to know?

The most important thing is to recognize the type of data one is dealing with and use an appropriate estimating technique. Qualitative dependent variables suggest using probit or logit; ordered qualitative dependent variables require ordered probit/logit; count data demand a Poisson model; duration data need a duration model; limited dependent variables point to Tobit estimation and selection models; time series data require analysis of unit roots and cointegration.

The most common tool used in applied work is undoubtedly the dummy variable. Practitioners need to be comfortable with the wide range of techniques that use this tool, such as estimating the influence of qualitative variables, structural break testing and estimation, and observation-specific dummy applications.

Econometric theory courses present a vast amount of information beyond these basics. Here are a dozen things practitioners should be sure to know.

1. *Instrumental variables.* Too commonly used not to know well.
2. *Mixed estimation.* Purists claim that incorporating stochastic information into estimation requires a Bayesian procedure, but mixed estimation is easier.

3. *Box–Cox*. An easy way to check functional form and test for logging data.
4. *Non-nested testing*. Easy to do and interpret.
5. *Bootstrapping*. Not easy to do, but important to understand. Many awkward testing problems can be solved via bootstrapping.
6. *Maximum likelihood*. Awkward estimation problems often require maximum likelihood estimation. Know how to find the likelihood function and how to get the computer to maximize it.
7. *ARIMA*. A fundamental benchmark for forecasting time series data.
8. *VAR*. A classic method for analyzing time series data. Foundation for the Johansen method for estimating cointegrating relationships.
9. *Heckman two-stage*. Know how to use this technique for correcting for sample selection bias; it is very popular, despite evidence indicating that MLE (and at times OLS) is superior.
10. *Identification*. Know how to check for identification, and realize that if an equation is not identified its estimation, using any technique, is without meaning.
11. *Panel data*. Understand the difference between fixed and random effects estimation and the circumstances in which one is more appropriate than the other.
12. *Nonstationarity*. What is it, why worry about it, how to test for unit roots, what is cointegration, what is the role of error correction models, and how to use software to estimate using the Johansen technique.

Finally, there is a variety of techniques not found in econometric theory texts, that to some define what is meant by applied econometrics. Examples are examining discrimination via the Blinder/Oaxaca decomposition, estimating consumer behavior by using AIDS, the "almost ideal demand system," estimating producer behavior by estimating a set of factor demand equations derived through duality from a translog cost function, knowing how to exploit "natural experiments," and understanding data filters and aggregation problems. An examination of these techniques lies beyond the scope of this book. (But see some limited commentary and references in the general notes.)

General Notes

22.1 Introduction

- Several texts have good expositions of how applied econometricians have examined classic topics. Berndt (1991), Thomas (1993), and Stewart (2004) are good examples. A crucial ingredient in applied econometrics is data; the *Journal of Economic Perspectives* has a section called "Data Watch," which brings data-related information to the attention of the profession. A similar section appears regularly in the features issues of the *Economic Journal*.
- Much of this chapter is based on Kennedy (2002) which contains a wide selection of quotes supporting the ten commandments and the critical views of applied econometrics instruction stated in the main body of this chapter. Some examples follow:

> My worry as an econometric theorist is not that there is tension between us (the theorists) and them (the applied economists). On the contrary, such tension can be healthy and inspiring. My worry is rather the lack of tension. There are two camps, a gap between them, and little communication. (Magnus, 1999, p. 60)

It is unfortunate that most textbooks concentrate on the easy estimation stage, when trouble is more likely to occur at the earlier specification stage. (Chatfield, 1991, p. 247)

At least 80 percent of the material in most of the existing textbooks in econometrics focuses purely on econometric techniques. By contrast, practicing econometricians typically spend 20 percent or less of their time and effort on econometric techniques per se; the remainder is spent on other aspects of the study, particularly on the construction of a relevant econometric model and the development of appropriate data before estimation and the interpretation of results after estimation. (Intriligator, Bodkin, and Hsiao, 1996, p. xiv)

Econometrics is much easier without data. (Verbeek, 2000, p. 1)

- Pagan (1999, p. 374) tells a story that captures neatly the difference between econometric theory and applied econometrics.

A Zen master presents his student with a stick and asks him what it is. The student responds with a description of its length and what it is made of, whereupon he is beaten with it. After a week of similarly precise answers and beatings, the student finally takes the stick and beats the master with it. As the student was meant to discover, it is not what you know about something which is important but rather how you use it.

- Heckman (2001) complains (p. 4) of "the current disconnect between economics and econometrics," notes (p. 3) that "in the past two decades, the gap between econometric theory and empirical practice has grown," and emphasizes (p. 4) that "command of statistical methods is only a part and sometimes a very small part, of what is required to do first-class empirical work."

22.2 The Ten Commandments of Applied Econometrics

- **Rule 1 Use common sense and economic theory.**

I was struck by how often I provided a service without doing anything that an academic researcher would recognize as statistics. Time and again I was thanked (and paid) for asking questions and suggesting perspectives that seemed to me to be little more than common sense. This highly developed common sense is an easily overlooked, but extraordinarily valuable commodity. (Trosset, 1998, p. 23)

Unfortunately, too many people like to do their statistical work as they say their prayers – merely substitute in a formula found in a highly respected book. (Hotelling *et al.*, 1948, p. 103)

- The role of theory extends beyond the development of the specification; it is crucial to the interpretation of the results and to identification of predictions from the empirical results that should be tested.

- **Rule 2 Avoid type III errors.**

… a Laurel and Hardy solution – where the initial question is transformed into an entirely different question and a solution offered. (Maddala, 1999, p. 768)

Far better an approximate answer to the *right* question, which is often vague, than an *exact* answer to the wrong question, which can always be made precise. (Tukey, 1962, pp. 13–14)

We found repeatedly that simple questions about seemingly minor details often bring to light misunderstandings of important issues. (Joiner, 1982, p. 333)

- **Rule 3 Know the context.**

Don't try to model without understanding the nonstatistical aspects of the real-life system you are trying to subject to statistical analysis. Statistical analysis done in ignorance of the subject matter is just that – ignorant statistical analysis. (Belsley and Welch, 1988, p. 447)

- Tweedie *et al.* (1998) and Pfannkuch and Wild (2000) provide examples of how a careful examination of the data-generating procedure has led to substantive insights. Burdekin and Burkett (1998) and Wilcox (1992) are examples of how not knowing the context can lead to error. Breuer and Wohar (1996) and Shannon and Kidd (2000) are examples in which knowing the institutional details of how the data were produced can aid an econometric analysis. Chatfield (1991) has some good examples of how empirical work can be greatly enhanced by being sensitive to the context of the problem and knowing a lot about one's data.

- **Rule 4 Inspect the data.**

  Every number is guilty unless proved innocent. (Rao, 1997, p. 152)

- Economists are often accused of never looking at their data – they seldom dirty their hands with primary data collection, using instead secondary data sources available in electronic form. Indeed, as noted by Reuter (1982, p. 137), "Economists are unique among social scientists in that they are trained only to analyze, not to collect, data. … One consequence is a lack of skepticism about the quality of data." Magnus (2002) cites Griliches as claiming that in economics poor data are blamed on the data collector, whereas in other disciplines the researcher him- or herself is taken to be responsible. Aigner (1988) stresses how dependent we are on data of unknown quality, generated by others for purposes that do not necessarily correspond with our own, and notes (p. 323) that "data generation is a dirty, time-consuming, expensive and non-glorious job." All this leads to an inexcusable lack of familiarity with the data, a source of many errors in econometric specification and analysis. This suggests that a possible route to finding better specifications is to focus on getting more and better data, and looking more carefully at these data, rather than on fancier techniques for dealing with existing data.
- EDA (exploratory data analysis) is an approach to statistics, introduced by Tukey (1977), which emphasizes that a researcher should always begin by looking carefully at the data in a variety of imaginative ways, such as via stem-and-leaf diagrams and box plots. Hartwig and Dearing (1979) is a good exposition; for examples see L.S. Mayer (1980) and Denby and Pregibon (1987). This approach cannot be recommended – it is evident that many statisticians (Ehrenberg, 1979), and especially econometricians, simply will not use the EDA techniques. But the spirit or "attitude" of EDA, as described by Cobb (1987, p. 329), is crucial:

  I find it useful to distinguish exploratory techniques such as stem-and-leaf diagrams and box plots, from exploratory attitudes: Does an author pay attention to such things as residuals, outliers, and the possible value of transforming? The former (techniques) are comparatively superficial, but the latter (residuals, outliers, transforming) lie close to the heart of data analysis.

- Maddala (1988, pp. 55–7) presents a nice example from Anscombe (1973) in which four sets of data give rise to almost identical regression coefficients, but very different graphs. Leamer (1994, p. xiii) has an amusing graph in which when graphed the data spell out HELP. Unwin (1992) discusses how interactive graphics should revolutionize statistical practice. Perhaps econometric software should have built into it some means of preventing a user from running a regression until the data have been examined! Tufte (1983) is a classic reference on how to display data visually. Hirschberg, Lu, and Lye (2005) is a tutorial on using graphs to look at cross-sectional data.
- Day and Liebowitz (1998) is a wonderful example of data cleaning. Maier (1999) is an excellent exposition of problems with data.

- **Rule 5 Keep it sensibly simple.**

  The general notes to sections 5.2 and 5.3 discuss the top-down versus bottom-up issue at some length, with related quotations.
- The conflict between simplicity and complexity arises in another context. Many econometricians employ the latest, most sophisticated econometric techniques, often because such techniques are novel and available, not because they are appropriate. Only when faced with obvious problems such as simultaneity or selection bias should more advanced techniques be employed, and then, as emphasized by Hamermesh (2000, p. 378), only

after a benefit–cost calculation has been applied, as he illustrates from his own work. Wilkinson and the Task Force on Statistical Inference (1999, p. 598) underline this view:

> Do not choose an analytic method to impress your readers or to deflect criticism. If the assumptions and strength of a simpler method are reasonable for your data and research problem, use it. Occam's razor applies to methods as well as to theories.

Maddala (1999, pp. 768–9) agrees:

> Think first why you are doing what you are doing before attacking the problem with all of the technical arsenal you have and churning out a paper that may be mathematically imposing but of limited practical use. Simplicity should be one's motto.

Cochrane (2001, p. 302) has an interesting perspective:

> Influential empirical work tells a story. The most efficient procedure does not seem to convince people if they cannot transparently see what stylized facts in the data drive the result.

- **Rule 7 Understand the costs and benefits of data mining.**

> ... any attempt to allow data to play a role in model specification ... amounted to data mining, which was the greatest sin any researcher could commit. (Mukherjee, White and Wuyts 1998, p. 30)
>
> ... data mining is misunderstood, and once it is properly understood, it is seen to be no sin at all. (Hoover, 1995, p. 243)

An extended discussion of data mining and the top-down versus bottom-up issue, with related quotations, can be found in the general notes to sections 5.2 and 5.3.

- Hand (1998) advocates the benefits of data mining. Kramer and Runde (1997) is an instructive example of the dangers of data mining. Sullivan, Timmermann, and White (2001) suggest a way of correcting for data mining when searching for calendar effects in stock market returns. Because extremely large data sets have become common, data mining has entered the mainstream, as evidenced by the introduction of the journal *Data Mining and Knowledge Recovery* and the development of data mining software, reviewed by Haughton *et al.* (2003).
- Testing procedures employed when data mining should be modified to minimize the costs of the data mining. Examples of such procedures are setting aside data for out-of-sample prediction tests, adjusting significance levels, and avoiding questionable criteria such as maximizing R^2. The Gets (general-to-specific) specification search software has these and many other sensible search procedures automated; references were provided in the general notes to section 5.2.

- **Rule 8 Be prepared to compromise.**

The Valavanis (1959, p. 83) quote from chapter 1 is worth repeating:

> Econometric theory is like an exquisitely balanced French recipe, spelling out precisely with how many turns to mix the sauce, how many carats of spice to add, and for how many milliseconds to bake the mixture at exactly 474 degrees of temperature. But when the statistical cook turns to raw materials, he finds that hearts of cactus fruit are unavailable, so he substitutes chunks of cantaloupe; where the recipe calls for vermicelli he uses shredded wheat; and he substitutes green garment dye for curry, ping-pong balls for turtle's eggs, and for Chalifougnac vintage 1883, a can of turpentine.

- **Rule 9 Do not confuse statistical significance with meaningful magnitude**

An extended discussion of statistical significance versus meaningful magnitude, with related quotations, is provided in the general notes to section 4.1. Meaningful magnitude is difficult to measure, and is usually measured through subjective evaluation of the context. A popular objective measure is an explanatory value's beta value – the number of standard deviations change in the dependent variable caused by a standard deviation change in the explanatory variable. This normalization tries to measure the impact of a "typical" change in the independent variable in terms of a "typical" change in the dependent variable.

- Another context in which unthinking significance testing can cause trouble occurs when testing down from a general to a specific specification. Adopting a traditional critical *t* value of 2.0 courts type II errors, namely omitting a relevant explanatory variable. It would be wiser to adopt a much smaller critical *t* value, say 1.0. For *F* tests the *p* value is an easier guide; rather than a critical *p* value of 0.05, a critical value of, say, 0.3 would be more suitable. This issue of the choice of a type I error was discussed earlier in the general notes to sections 4.1 and 5.2. Note how this is the opposite of worries about data mining, where the focus is on avoiding type I errors – including irrelevant explanatory variables.
- Fragility analysis, discussed in section 5.2 and its general notes, is a type of sensitivity analysis.
- Levine and Renelt (1992) is a notorious example of a sensitivity analysis. Abelson (1995) stresses that anticipation of criticism is fundamental to good research and data analysis.
- Welch (1986, p. 405) underlines the subjective/judgmental character of doing quality applied econometric work:

> Even with a vast arsenal of diagnostics, it is very hard to write down rules that can be used to guide a data analysis. So much is really subjective and subtle . . . . A great deal of what we teach in applied statistics is not written down, let alone in a form suitable for formal encoding. It is just simply "lore".

  Hendry (2000, chapter 20, 2001) does not fully agree with this; he claims that his Gets (general-to-specific) software does an amazingly good job of following rules to find appropriate specifications.

- The difficulty of teaching the lore of applied econometrics is highlighted in a well-known quote from Pagan (1987, p. 20):

> Few would deny that in the hands of the masters the methodologies perform impressively, but in the hands of their disciples it is all much less convincing.

- Although it is difficult for courses in applie econometrics to teach "lore," providing student with "experience" is more feasible. Of particula help in this regard are advances in computer tech nology that have lowered the cost of doing an teaching applied work, an increase in the num ber of journals providing access to the data use in their published articles, and websites full c data associated with textbooks. To provide use ful experience, however, instructors will have t design assignments that force students to fend fo themselves in real-world contexts, with vagu general instructions, rather than specific step-by step directions telling them what to do. Kenned (2002, reply) offers some examples.

22.3 Getting the Wrong Sign

- There is no definitive list of ways in whic "wrong" signs can be generated. In general, an theoretical oversight, specification error, dat problem, or inappropriate estimating techniqu could give rise to a "wrong" sign. This section based on Kennedy (2005), where references an additional examples can be found.
- Rao and Miller (1971, pp. 38–9) provide a example of how bad economic theory can lead a "wrong" sign. Suppose you are regressing th demand for Ceylonese tea on income, the pric of Ceylonese tea, and the price of Brazilian co fee. To your surprise you get a positive sign on th price of Ceylonese tea. This dilemma is resolv by recognizing that it is the price of other tea, su as Indian tea, that is the relevant substitute here
- The "*ceteris paribus* confusion" category fo "wrong" signs could be expanded to inclu examples that some might prefer to categorize foolishness on the part of the researcher.

  Reverse measure. Suppose you are regressi consumption on a consumer confidence measu among other variables, and unexpectedly obta a negative sign. This could happen because yo didn't realize that small numbers for the co sumer confidence measure correspond to hi consumer confidence.

  Common trend. A common trend could s what would otherwise be a negative relatio between two variables; omitting the common trend would give rise to the wrong sign.

Functional form approximation. Suppose you are running a 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) presents this example.

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, pp. 44–5). As another example, 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.

- A broader interpretation of "wrong" sign allows it to correspond to situations in which a statistically significant relationship is identified when no relationship may be present.

1. Suppose you have selected a set of firms with high profits-to-sales ratios and have regressed this measure against time, finding a negative relationship, that is, over time the average ratio declines. This result is likely due to the regression-to-the-mean phenomenon – the firms chosen probably had high ratios by chance, and in subsequent years reverted to a more normal ratio. As another example, 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. Both these examples illustrating the *regression-to-the-mean* phenomenon are exposited in Friedman (1992).
2. Omitting a relevant variable correlated with an irrelevant variable causes that irrelevant variable to proxy for the omitted variable and so appear to be relevant.
3. Regressing a random walk on an independent random walk should produce a slope coefficient insignificantly different from zero, but as seen in chapter 19, far too frequently does not. Spurious correlation associated with nonstationary variables is a source of "wrong" signs.
4. The Poisson model assumes that the variance of the number of counts is equal to its mean, whereas in reality there is typically overdispersion. Ignoring overdispersion causes the Poisson to produce unrealistically low standard errors, causing irrelevant variables to turn up "significant."

22.4 Common Mistakes

- The list of common mistakes is subjective, based on the author's experience refereeing empirical papers. It must be emphasized these mistakes are in addition to mistakes classified as ten commandments violations.

22.5 What do Practitioners Need to Know?

- Here are some more things that practitioners should know.

1. Give variables meaningful names; to facilitate interpretation, call the gender dummy

"male," if it is coded one for males, rather than calling it "sex."

2. Data for which percentage changes make more sense in the context of the problem you are investigating should be logged. Typically wages, income, price indices, and population figures should be logged, but age, years of education, and rates of change such as interest rates, should not be logged.
3. Recognize that bias is not sacred; allowing some bias can buy efficiency. Nor is efficiency sacred; forgoing some efficiency can buy robustness.
4. View a multicollinearity problem as equivalent to having a small sample. Realize that getting more information is the only solution.
5. Know how to analyze the size and direction of bias caused by an omitted variable.
6. Use a lagged value of the dependent variable to model dynamics and to proxy for unobserved factors whose omission would bias estimation.
7. Know how to estimate the percentage impact of a dummy variable on a dependent variable that has been logged for estimation.
8. To deduce the interpretation of a dummy variable specification, write out the specification for every category.
9. When testing for structural breaks, do not postulate a change in slope while fixing the intercept. Worry about whether it is legitimate to assume a constant variance across regimes.
10. When a theoretical relationship is manipulated algebraically to create an estimating equation, do not forget to apply the manipulations to the error term.
11. Know how to derive aggregate specifications from equations representing individual behavior.
12. Be aware that R^2 has no meaning if there is no intercept. In general, do not pay much heed to R^2.
13. Know how to check for outliers and influential observations. Know that these observations should be inspected, not automatically omitted.
14. Be familiar with the variety of diagnostic tests automatically printed out by econometric software, such as the RESET and DW tests.
15. Be familiar with BIC and AIC, and how they should be used.
16. Undertake predictive failure tests by using observation-specific dummies.
17. Know functional form options and their common uses.
18. Know that using OLS to estimate simultaneous equations should not necessarily be condemned.
19. Missing data should prompt the question "Why are they missing?" If there are no selection problems, if some but not all regressor values are missing for an observation, replacing missing explanatory variable data with estimated values should be considered, rather than omitting that observation.
20. Be aware that the "best" forecast is a combined forecast, and should be evaluated using a context-specific loss function.
21. Testing for exclusion of independent variables should adopt a low critical t value (1.0 or less, e.g., rather than the traditional 2.0) to minimize the influence of type II errors (i.e., to avoid omitting a relevant variable). In general, pretesting of any kind should be conducted using a significance level much higher (say, 25%) than the traditional 5%.
22. Robust estimation can play an important role in sensitivity analysis – check if parameter estimates change much when a robust estimation procedure is used.
23. An insignificant DW statistic with cross-sectional data should not be interpreted as indicating lack of error autocorrelation. A significant DW statistic should not be ignored; it probably reflects a nonlinearity with the data ordered on the basis of an explanatory variable.
24. If you are estimating the impact of a policy simulate a typical policy change to see if the estimated results are reasonable.

- A good exposition of the Blinder/Oaxaca methodology can be found in Berndt (1991, pp. 182–4). Oaxaca and Ransom (1994) discuss means of breaking the discrimination portion of the difference between blacks and whites into an advantage to whites and a disadvantage to blacks. Couch and Daly (2002) discuss another extension of this methodology in which the residual is decomposed into a portion reflecting a movement of blacks up the distribution of white residuals, and a change in inequality as reflected by a decrease in the residual variance. Fairlie (2005) explains how the Blinder/Oaxaca method can be extended to probit/logit estimation. A problem with the Blinder/Oaxaca methodology is that the results are sensitive to the choice of reference category when using dummy variables; see Gardeazabal and Ugidos (2004) for an explanation and a means of dealing with this problem.
- The classic reference for AIDS is Deaton and Muelbauer (1980). Alston, Foster, and Green (1994), Moschini (1998), Wan (1998), and Buse and Chen (2000) discuss variants of this model, and offer several practical suggestions for estimation. Pollak and Wales (1992) is a general reference on modeling and estimating consumer demand systems. Fisher, Fleissig, and Serletis (2001) compare flexible demand system functional forms. Deaton (1997) is a good reference for the analysis of household survey data. Keuzenkamp and Barton (1995) is an instructive history of testing the homogeneity condition in consumer demand.
- Berndt (1991, chapter 9) is a good presentation of how duality is used to help estimate cost and production functions. Burgess (1975) is a good summary of this technique, with a critical discussion based on the fact that the translog function is not self-dual. Coelli, Rao, and Battese (1998) is an excellent exposition of efficiency and productivity analysis, with introductory discussions of factor productivity indices, data envelopment analysis (DEA), production and cost function estimation, and stochastic frontier production function estimation. Factor productivity indices are constructed as weighted averages of outputs divided by weighted averages of inputs, with a variety of competing formulas used for the weights. DEA is a means of estimating a production possibilities frontier by using principles from linear programming. Stochastic frontier analysis estimates a production possibilities frontier by incorporating a requirement that error terms must be negative. Kalirajan and Shand (1999) and Kumbhakar and Lovell (2000) are more advanced references on stochastic frontier analysis.
- Baxter and King (1999) discuss ways of detrending data to identify business cycles, such as by using the popular Hodrick–Prescott filter.
- Aggregating data often causes substantive econometric problems. McGuckin and Stiroh (2002) have a good discussion of this issue.
- Much recent empirical work in economics, particularly that associated with policy issues, has attempted to find and exploit natural experiments, or to construct reasonable facsimiles thereof. Krueger (2001, p. 244) describes this change in method as follows. "The empirical work that was common in the 1970s was designed largely to derive parameter estimates as input for a particular theory, and empirical tests were highly dependent on theoretical assumptions. Today it is more common to search for natural experiments that would provide persuasive evidence either refuting or supporting a hypothesis."

  A natural experiment is a situation in which some feature (often unintended) produces exogenous variation in what would otherwise be an endogenous variable, allowing the researcher to estimate the impact of a treatment. This could come about, for example, if changes in social service benefits affected some groups but not others, or if one state changed its minimum wage but another did not. Less obviously, it may be possible to identify an instrumental variable which can capture this exogenous variation. Meyer (1995) and Blundell and Costa Dias (2000) discuss several methods of evaluating the impact of a change/treatment.

  1. A socialized experiment, in which a set of randomly drawn people experiences the treatment. The impact of the treatment is measured

by the change experienced by this group. Greenberg, Shroder, and Onstott. (1999) is a good survey of social experiments.

2. A "natural experiment," in which one group has experienced the treatment, whereas another, comparable group has not. The impact of the treatment is estimated by looking at the difference between the changes experienced by the two groups before and after the treatment. This method is sometimes referred to as the "difference in differences" method. Murray (2006a, pp. 656–64) has a good exposition, along with several examples (pp. 671–6) in the form of exercises.
3. The "matching" method, in which an artificial control group is selected from among those not experiencing the policy. Every individual chosen to be in the control group is chosen on the basis that his or her "propensity score," the estimated probability of an individual being in the experimental group, matches that of an individual in the experimental group.
4. A selection model is estimated, using, for example, the Heckman two-stage estimation procedure, to avoid selection bias caused by people self-selecting themselves into the treatment group.
5. An instrumental variable estimation method is employed, to circumvent contemporaneous correlation between the treatment and the error caused by omission of an explanatory variable.

There is some evidence that the second and third of these methods do not work well in practice, suggesting that nonexperimental procedures do not produce reliable estimates; see Fraker and Maynard (1987), LaLonde (1986), LaLonde and Maynard (1987), and Bertrand, Duflo, and Mullainathan. (2004). See also Friedlander, Greenberg, and Robins. (1997) and Dehjia and Wahba (2002). The February 2004 issue of the *Review of Economics and Statistics* contains a symposium on this subject; Michalopoulos, Bloom, and Hill (2004) and Agodiri and Dynarski (2004) repeat the message that matching methods do not work well.

The instrumental variables approach is called a "natural experiment" methodology because it is applicable whenever the nonexperimental data generation process has inadvertently created suitable instrumental variables. For example, attempts to estimate the returns to education are frustrated by the fact that an unmeasured variable "ability" affects both earnings and the number of years of education chosen. This would create upward bias to the estimate of returns from schooling. Measurement error in the schooling measure, another source of contemporaneous correlation between the error and the regressor schooling, would bias this estimate downward. A third source of this contemporaneous correlation is that the returns to schooling might vary across individuals, being higher for low-schooling people. This would bias the returns from schooling measure downward. These biases could be avoided by finding a variable that affects schooling but not earnings, and so could serve as an instrument. Several ingenious suggestions have been proposed, such as the quarter of the individual's birth and distance to college. Card (2001) has a good survey. The instrumental variable approach solves the omitted variables problem by estimating using only part of the variability in the explanatory variable (a part uncorrelated with the error, as explained in chapter 9). Angrist and Krueger (2001) and Wooldridge (2002, chapter 18) are good discussions. Stock and Watson (2007, chapter 13) and Murray (2006a, chapter 15) have good textbook discussions of experiments and quasi-experiments in econometrics.

- The natural experiment literature has given rise to different interpretations of the impact of a "treatment," discussed in the general notes to section 9.3. The *average treatment effect* (ATE) is the expected effect of the treatment on a randomly drawn person from the population eligible for treatment. The *average treatment effect on the treated*, ATET, is the expected effect on those who actually were treated. The *local average treatment effect* (LATE) is the expected effect of the treatment on those whose behavior is captured by the instrumental variable used for estimation. An interesting problem here is how to measure an ATE if the treatment affects different individuals differently. This could happen if a nonlinearity is involved (in a logit model, for example, the impact of the treatment depends on the index value of the individual) or if an interaction term is involved (so that, for example, the treatment effect varies with income). One way of measuring ATE is to find the ATE for an artificial individual with the average characteristics of the sample. (Such a person might, for example, be sixty percent female if sixty percent of the sample is female.) A competing way is to find the ATEs for all individuals in the sample and average them. These two measures will not be the same, and in both cases rely on the sample being representative of the relevant population. A third way to report here is not to report ATE, but rather report the impact on several different types of individuals. Probably the best thing for a researcher to do is report in all three ways.

Technical Notes

22.2 The Ten Commandments of Applied Econometrics

- Are data-miners made, or born? Green (1990) gave 199 students the same data but with different errors, and asked them to find an appropriate specification. All students had been taught that models should be specified in a theoretically sensible fashion, but some were also taught about how to use F tests and goodness-of-fit for this purpose. These latter students were quick to abandon common sense, perhaps because using clearly defined rules and procedures is so attractive when faced with finding a specification.

22.5 What do Practitioners Need to Know?

- Earlier lists referred to things practitioners needed to know well enough to be able to apply. There are some technical things, however, that should be understood well enough to avoid being perplexed by others' references to them. Appendix B, "All about Variance," covers a specific set of technical material of this nature, for example. Here are some additional examples.

1. *Mechanical properties of OLS* Knowing these properties, such as that the sum of residuals is zero, can frequently be useful.
2. *F versus chi-square tests* Asymptotically, an F statistic when multiplied by its numerator degrees of freedom (the number of restrictions under test) is equal to its corresponding chi-square test. In small samples it is not known which test is better, but a tradition of using the chi-square test has developed.
3. *Bayesian approach* Bayesian estimation has not yet reached the stage where practitioners need to be able to do it. But everyone should know how this estimation procedure differs – that Bayesian estimates incorporate prior beliefs explicitly, are weighted averages of estimates associated with different values of nuisance parameters, and pay attention to the purpose of the analysis, through a loss function.
4. *GMM* Generalized method of moments is a unifying view of econometric estimation that in theory is impressive but in practice has not lived up to its promise.
5. *LM, LR, and W* These test statistics are asymptotically equivalent, the choice among them typically made on the basis of computational convenience. LR is judged to be best in small samples.
6. *Lag operator* Algebraic manipulation of theoretical relationships to produce an estimating equation is dramatically simplified by using the lag operator, which, amazingly, can be treated as an algebraic term.
7. *Nonparametrics* Computer software has not yet reached the stage at which practitioners will use nonparametric estimators, but the essence of these estimation techniques, and their need for large sample sizes, should be understood.
8. *IIA* Lack of independence of irrelevant alternatives causes problems for multinomial logit.

9. *Condition index* The condition index is judged to be the best indicator of multicollinearity, but software typically does not calculate it easily.

- Derivation of the Blinder/Oaxaca method utilizes the result that the OLS estimating line passes through the average of the observations, or equivalently, that the sum of the residuals equal zero. A consequence is that the average of the estimated dependent variable observations is exactly equal to the OLS estimating equation evaluated at the average values of the explanatory variables.

Chapter 23

Computational Considerations

23.1 Introduction

In the early days of econometrics, calculating estimators was a major task. Hours would be spent using a desktop calculator (calculator, not computer!) to compute sums of products leading eventually to the ordinary least squares (OLS) estimate. Students were taught computational tricks to facilitate these calculations. For example, when using OLS, if there are two explanatory variables a 3×3 matrix must be inverted (the $X'X$ matrix). But if the means are removed from the data this 3×3 matrix shrinks to a 2×2 matrix (because the row and column associated with the constant term disappear).

The invention of the computer did away with this drudgery. OLS estimates could be found simply by submitting to a computer center a program coded on punched cards, with output returned quickly, depending on the length of the queue. But more complicated estimation problems, rarely attempted by hand, were challenges even for the early computers. In the mid-1960s, for example, estimation of a three-equation simultaneous equation system using full-information maximum likelihood (FIML) would take several hours of central processing unit (CPU) time, and a university computer center would only agree to run this calculation overnight on the weekend. Make one mistake in your coding and you would have to wait a week for another run!

The advent of the desktop computer, the development of user-friendly econometrics software, and the spectacular increase in computer power over the past few years has dramatically altered the practice of econometrics. There are several major changes:

1. Many standard estimation procedures, such as OLS, instrumental variable (IV), and maximum likelihood for a wide variety of standard cases (probit, logit, Tobit, FIML, count data, Box–Cox, and sample selection, for example), can be estimated virtually instantaneously at the push of a software button. This has led to a huge increase in data mining, as researchers search through various options to find results they prefer.