Title

A statistical symphony: Instrumental variables reveal causality and control measurement error

Permalink

https://escholarship.org/uc/item/5v21m05h

ISBN

978-0-19-967254-7

Author

Kendall, Bruce E.

Publication Date

2015-01-29

Peer reviewed

A statistical symphony: Instrumental variables reveal causality and control measurement error

Bruce E. Kendall

Published as chapter 7 of G.A. Fox, S. Negrete-Yankelevich, and V.J. Sosa, eds. *Ecological Statistics: Contemporary Theory and Application* (2015: Oxford University Press).

7.1 Introduction

The goal of much ecological research is to determine causality, whether in test of ecological theory or to develop predictive power for making management decisions in the face of environmental change. In its simplest form, we want to test whether, and how strongly, x causes y — for example, does lizard density determine the density of spiders, one of the lizard's prey items (Spiller & Schoener 1998)? Although it is straightforward to demonstrate whether there is correlation between x and y, that alone does not tell us whether the observed relationship is actually an epiphenomenon of both variables being caused by a third factor, z (e.g., moisture; Spiller & Schoener 1995). The reason that experiments are so powerful is that, when done well, they rule out alternate causal explanations of the observed relationships. *Randomization* ensures that variation across sample units in z (and p and q and however many other factors might be relevant) is assigned randomly with respect to our experimental variable, x. Although we still can't attribute the response of a single sample unit to its value of x (it may, after all, have an extreme value of z), this ensures that our estimate of the statistical relationship between y and x has the desirable property of **consistency**. In brief, a consistent statistical estimator converges to the quantity being estimated—a clear prerequisite to valid statistical inference. Inconsistency feels a lot like bias in that it leads to systematically incorrect parameter estimates and P-values. However, unlike bias, which attenuates with sufficiently large sample size, inconsistency is structural, and cannot be remedied simply by collecting more data.

However, by their very nature experiments are limited in spatial and temporal extent, and abstract away much of the complexity of real ecological systems. To demonstrate causality in the field, ecologists have made use of *natural experiments*, in which natural variability in x across sites (e.g., differences in lizard density across islands) is used as a substitute for an experimental treatment. The problem, of course, is that a lot of other stuff besides x may be varying across sites. The usual prescription is to make sure that the sites are as similar as possible in every aspect other than x, but this is limited to the variables we can measure, and will be difficult to maintain across more than a few variables anyway. Furthermore, this isn't the right analogy to controlled experiments. True experiments don't "hold everything else equal;" they hold some things equal, but randomize to account for everything else. Natural experiments, even if carefully matched across a set of potential confounding variables, lack this randomization, so we cannot be sure that the observed variation in y is caused by x or by an unobserved z. We have seemingly lost our ability to make strong causal conclusions, which leads many ecologists to discount natural experiments, chanting the mantra "correlation does not imply causation." This view was reinforced by a number of influential books in the latter part of the 20th century (e.g., Underwood 1997), and the standard biometry text books focus heavily on analysis of experimental data, maintaining and artificial distinction between ANOVA and linear regression (see Chapter 1 [Introduction]).

In many cases we want to do more than simply determine the presence or absence of a causal relationship (the usual outcome of and ANOVA), and estimate the the magnitude and direction of the relationship (information provided by the parameter estimate in a regression). If the unobserved variable z is correlated with x, the predictor variable of interest, then, even if there is a strong causal relationship between x and y, this relationship *cannot* be estimated from the simple statistical association between x and z (Figure 7.1).
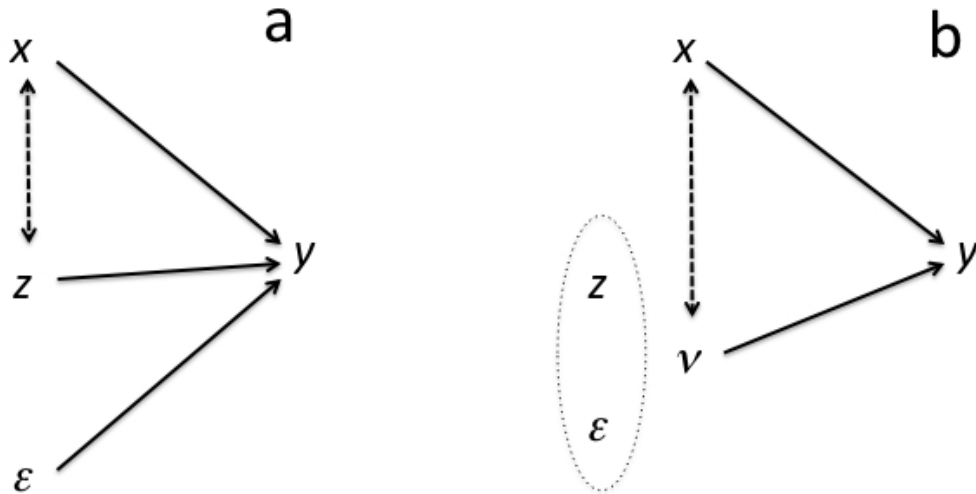
Figure 1: *Figure 7.1:* The omitted variable problem. (a) Suppose that y is causally influenced by both x and z (solid arrows), as well as host of small effects that are collected into the residuals ϵ. The dominant causal variables x and z are correlated (dashed arrow); this could result from direct or indirect causation in one or both directions, or from statistical non-independence in a finite sample (e.g., x and z have similar patterns of spatial autocorrelation). As long as the residuals are uncorrelated with x and z, then ordinary least squares (OLS) regression (e.g., `lm(y ~ x + z)`) will provide consistent estimates of the coefficients relating each of the predictor variables to the response variable. Each coefficient will be a partial regression coefficient, which accounts for the effects of the other predictor variable — this accurately represents the direct causal effect of each variable on its own. The only effect of the correlation between x and z is to (possibly) inflate the uncertainty around the estimates. (b) If z is unobserved, then the residuals from the *causal* relationship between x and y (not to be confused with the statistical relationship that we see in a bidirectional scatterplot) are $\nu = z + \epsilon$, which are correlated with x. By construction, OLS regression finds parameter estimates that *eliminate* any such correlation; these estimates may be consistently larger or smaller than the causal relationship, and may not even have the correct sign!

Economists have had to face this problem head-on, because most of the questions they are interested in are not amenable to controlled experiments, and they aim to inform decisions in the real world outside the library. Thus most of their tests of causality have to make use of natural experiments, and they have developed techniques that allow them to estimate the causal relationship between x and y in the face of uncontrolled variation in z. One approach that is general and powerful, and almost completely missing from the ecologist's toolbox, is the method of **instrumental variables** (IV), which is the subject of this chapter. An instrumental variable (or more commonly, a collection of instrumental variables) can effectively randomize x with respect to z, allowing us to infer the causal effect of x on y. The actual statistical implementation of this approach is only marginally more complex than least squares regression, and the whole thing can feel a bit magical. The "trick" is that instrumental variables need to meet some very rigorous requirements, so that finding a measurable IV (or even worse, identifying an IV from among a set of measurements that have already been taken) is challenging and may not always be possible.

Before proceeding further let's take a look at the impact such an unmeasured variable has on a regression equation.

7.2 Endogeneity and its consequences

From a formal statistical perspective, **endogeneity** is the presence of a non-zero covariance between one or more predictor variables and the residuals of the model (the name comes from the fact that one cause of such covariance is bi-directional causality, in which the predictor variable is not independent of the response variable). To see how this can arise, suppose that spider density (y) depends on lizard density (x) and moisture (z): $y \sim x + z$. Furthermore, suppose that moisture also affects lizard density, so that $cov(x, z) \neq 0$. The full model that we want to estimate is

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \epsilon_i,$$

where β_1 is the marginal effect of x on y — that is, a unit increase in x, holding all else equal, will increase y by β_1. This is illustrated by hypothetical data in Figure 7.2, where, for ease of visualization, "moisture" (z) takes on just two values. The two predictor variables have opposite effects on y, and are correlated with each other. If we have measurements of all three variables, we can use multiple regression to get consistent parameter estimates.

However, if we haven't measured z, we are left with trying to estimate a short-form model,

$$y_i = \beta_0 + \beta_1 x_i + \nu_i,$$

where $\nu_i = \beta_2 z_i + \epsilon_i$; this is the black line in Figure 7.1 (this makes clear that the "error" term in a regression includes the causal effects of unmodeled variables). The problem is, the residuals from this model are not independent from x: $cov(x, \nu) = \beta_2 cov(x, z)$. In Figure 7.1, the ν_i are the residuals from the black line to the points; they are mostly negative for small values of x and mostly positive for large values of x. If we ignore this and fit the short model using OLS, the estimated coefficient of x will be a combination of the direct effects of x on y and the indirect effects via z, giving a slope estimate that bears little relationship to the underlying causal effect (Figure 7.1). This effect is not simply bias, which could be reduced by increasing the sample size. Even in the limit of infinite data, this estimate will not converge on β_1: in the presence of correlations between the residuals and the predictors, OLS parameter estimates are inconsistent. For the data in Figure 7.1, the short-form regression gives the wrong sign for the effect of x on y; this will not always be the case, as it depends on the correlation structure among the variables. Nevertheless, even if by good luck the regression estimates the correct sign of the relationship, the confidence intervals and P values will be meaningless.

In general, we will not be able to identify the existence of the missing variable z by looking at diagnostic plots of the regression of y on x. In the illustrated example, with a single predictor variable and a single omitted variable, it appears that the skew of the estimated residuals varies as a function of x, which might be revealed as a trend in the conditional median (not mean) of the residuals. However, this is unlikely to generalize to the more typical situation of multiple predictor and omitted variables. Thus, we must use our theoretical
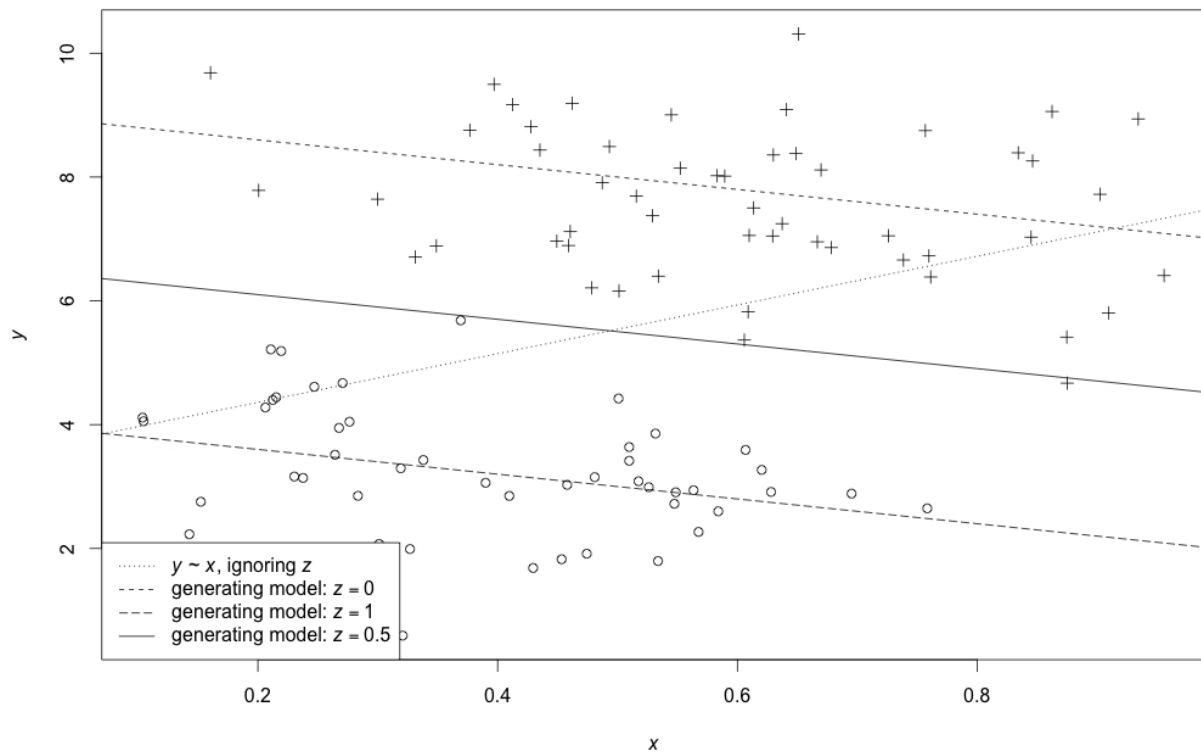
Figure 2: *Figure 7.2:* Omitted variables lead to inconsistent estimates. These simulated data were generated by $y_i = 4 - 2x_i + 5z_i + \epsilon_i$, where z takes on values of zero (circles) or one (plusses) and is correlated with x. The dashed lines show the generating model with z taking on values of zero and one; the solid line is the marginal effect of x evaluated at the "average" value of z (which is ultimately the effect that we are trying to recover). However, regressing y on x gives a slope estimate of 4.12 (dotted line), which is not just the wrong sign but very convincingly so: $P = 0.001$ for the t-test against a null hypothesis of zero slope. Code for generating and analyzing these data is in the appendix.

understanding of the system generate lists of putative variables that might causally influence the response variable and could be correlated with the predictor variable. If we have measured these variables then we can add them to the model; but if we are not so fortunate, then we cannot know whether our analysis is science or science fiction.

7.2.1 Sources of endogeneity

The endogeneity in the example of the previous section is generated by an *omitted variable* (confusingly, the phenomenon is conventionally called "omitted variable bias," even though the problem is inconsistency rather than bias). Given the complexity of ecological causality, it is likely that this is a common phenomenon in ecological data. Without measurements of the variable in question, or experimental approaches that randomize the observations with respect to the omitted variable, it is difficult to imagine how to address this problem. We shall see in the next section that instrumental variables, if they can be found, can control for this effect, allowing us to draw causal conclusions. There are two additional common causes of edogeneity: measurement error and bidirectional causality.

Measurement error in an predictor variable also generates endogeneity. Suppose that

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

but that we have an imperfect observation of x,

$$x_i^* = x_i + \nu_i.$$

We can estimate a model with the observed values,

$$y_i = \beta_0 + \beta_1^* x_i^* + \epsilon_i^*,$$

where $\epsilon_i^* = \epsilon_i - \beta_1 \nu_i$.

Typically we expect the measurement error ν_i to be uncorrelated with the true value x_i. However, this means that, assuming that the expected value of ν is zero,

$$\begin{aligned} \mathrm{Cov}(\epsilon^*, x^*) &= \mathrm{Cov}(\epsilon - \beta_1 \nu, x + \nu) \\ &= \mathrm{Cov}(\epsilon, x) + \mathrm{Cov}(\epsilon, \nu) - \beta_1 \mathrm{Cov}(\nu, x) - \beta_1 \mathrm{Cov}(\nu, \nu) \\ &= 0 + 0 - 0 - \beta_1 \mathrm{var}(\nu). \end{aligned}$$

Thus the residuals between the actual causal relationship and the observed data are correlated with the observed predictor x^*, with sign opposite to the sign of the underlying relationship β_1. The former generates inconsistent estimates; the latter ensures that the inconsistency is in the direction of zero. Because this is inconsistency rather than bias, it cannot be reduced simply by increasing the number of data points in the regression (once again, the conventional terminology — "measurement error bias" — muddies this distinction). This is why precise measurement of the predictor variable is one of the key assumptions of the linear model. However, this assumption is rarely discussed in ecological studies that use linear models.

This effect is not entirely unknown to ecologists (e.g., Solow 1998, Freckleton 2011). In general, estimating and correcting the magnitude of this "attenuation" requires information about the measurement error variance. If the estimates of x come from a well-characterized statistical sampling process, this variance is available and there are standard techniques for correcting the attenuation (Buonaccorsi 2010). Absent such information, however, there is little that can be done to eliminate the endogeneity directly (although there are specialized tools for certain circumstances, such as using state space models to estimate parameters of autoregressive processes). However, recognizing that measurement error creates this inconsistency by inducing endogeneity shows that we can use apply tools that address endogeneity directly.

The third major source of endogeneity is *bidirectional causality* (often called "simultaneity" in the economics literature), in which two variables causally affect each other. For example, Bonds et al. (2012) sought to analyze whether a county's biodiversity affects the disease burden on the its human inhabitants. An important

control variable for this analysis is the country's per capita income: all else equal, we expect richer countries to have lower disease burdens, because of investment in public health and medical facilities. However, a high disease burden tends to reduce per capita income, by reducing productivity and life expectancy. Clearly, then, per capita income cannot be treated as "independent" in a regression in which disease burden is the dependent variable. The solution is to use a "simultaneous equation model" (not to be confused with a *structural* equation model; chapter 8), in which two equations, one with disease burden as the predictor variable and the other with per capita income as the dependent variable, are solved simultaneously. There is still endogeneity within each equation, however, which we will address below.

Interestingly, endogeneity can arise even in randomized experiments, if the putative causal variable isn't under direct experimental control. For example, suppose you are using microcosms to study the effect of phytoplankton density on *Daphnia* growth. You manipulate phytoplankton density indirectly by applying one of a number of different nutrient supplies to each microcosm. You know that temperatures are uneven in your growth chamber, so you use a randomized block design to ensure that treatments are random with respect to these position effects, and then you regress the observed *Daphnia* growth rate on the observed phytoplankton density (measured as the opacity of the water in the container). However, there may be a problem: the phytoplankton in the microcosms in the "warm" part of the chamber will probably grow faster than average and achieve a higher than average density, given the nutrient supply level; and the *Daphnia* may also grow faster than average, given the phytoplankton density. Thus *Daphnia* in a warm location have a double advantage: they get more food than average, given the experimental treatment; and they grow faster than average, given the amount of food. The unobserved temperature variation is affecting both the causal variable and the response variable, creating endogeneity in a way that is formally identical to the effect of individual quality on the life-history tradeoff analysis; and because it affects both sides of the regression equation, the careful experimental randomization doesn't help. In this case, with a positive relationship between *Daphnia* growth and phytoplankton density, and a positive correlation between the conditional responses of phytoplankton and *Daphnia* to random variation, the regression will overestimate the magnitude (and probably, the statistical significance) of the relationship. You could analyze the response of *Daphnia* to the level of nutrient addition; but while this gives a statistically consistent result, it probably will not be biologically useful.

7.2.2 Effects of endogeneity propagate to other variables

By now it should be clear that endogeneity leads to serious problems in identifying and estimating causal relationships in bivariate data. Even more distressing, however, is that when performing multiple regression (more than one predictor variable), a single endogenous variable can cause the estimated coefficients of the other predictor variables to be inconsistent as well. This will occur in the common situation where the predictor variables are correlated with one another. Not surprisingly, the effect becomes stronger as the correlations among the variables becomes stronger. Thus, one badly endogenous variable can spoil a whole regression! In particular, because measurement error is a source of endogeneity, this means that one should think very carefully before introducing a poorly measured variable into a model otherwise comprising accurately measured quantities.

7.3 The solution: instrumental variable regression

Instrumental variable regression solves the endogeneity problem by using a suitable new variable to "instrument"[1] the endogenous predictor variable, controlling for the correlation between the predictor variable and the residuals. Mathematically, this solution is disarmingly simple; the challenge is finding a variable that meets the conditions to be an instrument.

---

[1]The conceptual origin of the term "instrumental variable" is not recorded; the term was introduced without comment in a 1945 dissertation (Morgan 1990). However, insofar as the instrumental variable is (imprecisely) measuring the variation in the regressor that is not correlated with the error term, the term may have it roots in the idea of a scientific instrument as a measurement device. Of interest to biologists, the technique itself was first developed 20 years previously by Sewall Wright's father, probably with Sewall's input (Stock and Trebbi 2003).

An instrumental variable must satisfy two conditions (Figure 7.3):

1. *Relevance:* It must be correlated with the endogenous regressor variable; the stronger the correlation, the more effective it will be.
2. *Exogeneity:* It must have no direct causal effect on the response variable, and be correlated with the response variable *only* because of its relationship to the endogenous regressor variable. This is often called the *exclusion condition.*

Finding a variable that satisfies the first criterion is usually easy. The second condition, which has the effect that the instrumental variable is uncorrelated with the residuals of the regression, is the bottleneck. In economics, uncovering a new instrument for an important class of problems can launch a successful academic career. I discuss possible examples of instrumental variables in ecology below; first, let us examine how instrumental variables solve the endogeneity problem.
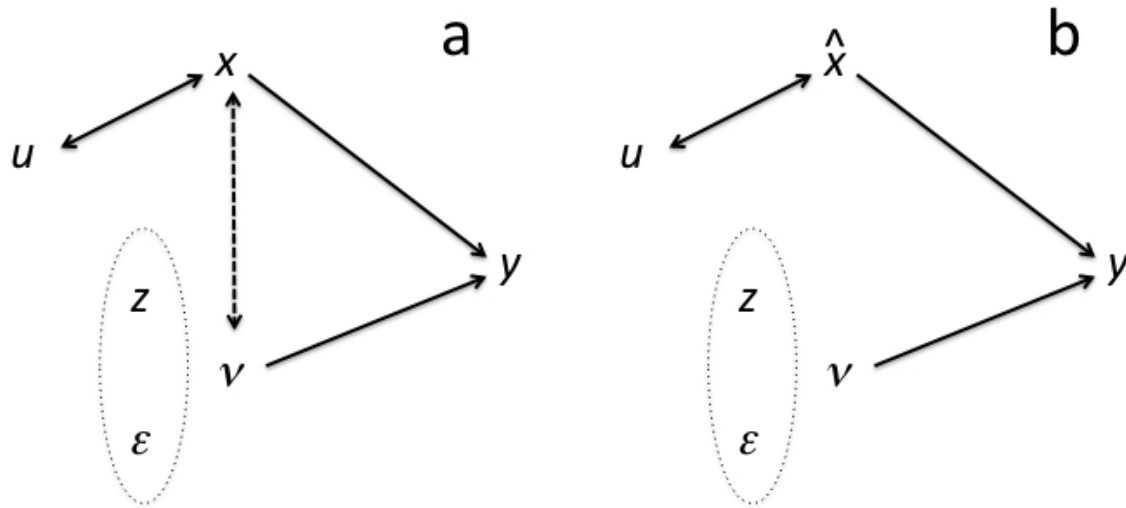

Figure 3: *Figure 7.3:* The instrumental variable solution to the omitted variable problem of Figure 7.1. (a) The instrumental variable, u, has a causal or structural relationship with the predictor variable x, which creates a correlation between the two variables. In addition, u has no relationship with the response variable, y, except through the pathway involving x; in particular, this means that u is uncorrelated with z, ϵ, and ν. In essence, u captures (some of) the variation in x that is uncorrelated with z and ν. (b) Thus, the predicted value of x given u, $\hat{x}$, is uncorrelated with ν and hence does not suffer from endogeneity. As a consequence, the OLS regression of y on $\hat{x}$ gives a consistent estimate of the causal relationship between x and y; the cost is greater uncertainty in the parameter estimates.

The general idea is that the instrument represents a treatment that, while not experimentally randomized, is effectively random with respect to the process that introduces endogeneity. As a concrete example, a central feature of life history theory is the notion of a tradeoff between current and future reproduction (De Jong & Van Noordwijk 1992). Consideration of energetics suggests that, for a given individual, a greater investment in reproduction today will reduce the individuals resources for survival and future reproduction; evolved reproductive strategies are thought to take this into account. When trying to measure this tradeoff in the field, it is tempting to simply look for a relationship between a given years reproduction and some measure of subsequent reproduction (survival, reproductive success in the following year, etc.). However, doing this as simple regression often results in positive estimates of the coefficient for the effect of current reproduction on future success. This occurs because the analysis looks across individuals instead of within individuals, and any variation in individual quality (i.e., some individuals, whether because of good genes, good parents, or good environments, are better than average at both current and future reproduction) tends to swamp the effects of within-individual tradeoffs. This unmeasured quality variable is like the omitted variable in Figure 7.1. Formally, what we are interested in is the coefficient β_1 in

$$F_i = \beta_0 + \beta_1 C_i + \beta_2 Q_i + \epsilon_i,$$

where F is future reproductive success, C is current reproductive success, and Q is quality, which in addition to affecting future reproductive success as shown in the equation, is correlated with current reproductive success,

$$C_i = \gamma_0 + \gamma_1 Q_i + \delta_i.$$

If Q is not observed, then running the regression $F \sim C$ will be analogous to fitting equation 7.2, and will clearly give the wrong estimate for β_1.

What we need is a way to manipulate current reproduction independently of quality. Thinking specifically of birds, we can imagine that nest predators might do this for us. We have to think carefully about this, because it might be that high quality individuals (from the perspective of potential reproductive success) might also be better at defending against predators, but in at least some cases we could be justified in concluding that this correlation is weak at best (a rigorous argument would entail a detailed consideration of antipredator behavior, nest siting, etc.). Now consider circumstances in which predators consume all eggs before they hatch, and the pair doesnt renest. At this point the males energetic investment is modest, and we might expect that males who have had their eggs depredated should have on average, more resources for future reproduction than those who havent. If, on average, birds whose nests were depredated have the same quality as those whose nests were not depredated, then a regression of future success on depredation status should not suffer from endogeneity, and so the resulting parameter estimate is consistent.

However, the relationship between depredation status and future success is not what we are interested in. Define depredation status to be instrumental variable, Z. As an IV, it must be correlated with C but have no direct causal relationship with F. Another way of saying the latter is that adding Z to the full model does not improve the fit of the model. A third way of saying it is that Z is uncorrelated with both Q and with ϵ; this implies that Z is also uncorrelated with the residuals of $F \sim C$.

To get an unbiased estimate of the parameter describing the effect of current reproduction on future success, conditioned on quality, then we combine the relationship between depredation and future success with the relationship between current reproduction and depredation. First we estimate the relationship between the predictor variable (current reproduction) and the instrumental variable (depredation status):

$$C_i = \alpha_0 + \alpha_1 Z_i + \omega_i.$$

This is called the 'first-stage regression.' Then we estimate the model we conceptually described when we introduced the depredation variable,

$$F_i = a + bZ_i + e_i$$

This is the 'reduced form regression.' Through some rather tedious statistical theory, it can be shown that $\rho = b/\alpha_1$ is a consistent estimate of β_1 (e.g., Greene 2011).

Although the above calculation is straightforward for that simple example, it is rather more complex when there are multiple predictor and/or instrumental variables. Instead this analysis is commonly done using a

procedure called *two stage least squares* (2SLS), which scales to models of arbitrary complexity. It starts with the first stage regression, but the creates a synthetic variable that is, in effect, the predictor variable scrubbed of its endogeneity. First, estimate the coefficients from the first stage regression, call them $\hat{\alpha}_0$ and $\hat{\alpha}_1$. Then, use these to estimate the expected value of C given Z:

$$\tilde{C}_i = \hat{\alpha}_0 + \hat{\alpha}_1 Z_i.$$

$\tilde{C}$ is the conditional expected value of current reproduction, conditioned on depredation status. Then regress F on the $\tilde{C}$ (the 'second stage regression'); the resulting regression coefficient for $\tilde{C}$ is also a consistent estimate of β_1 (for the simple model here, it is identical to ρ).

To develop a graphical intuition, we return to the simulated data of Figure 7.2. Figure 7.4(a) shows the first stage regression, the relationship between the predictor variable (x) and an instrumental variable (u). Notice that, whereas u is uncorrelated with z, the residuals from the first stage regression are almost perfectly correlated with the omitted variable, z (positive residuals for $z = 1$; negative residuals for $z = 0$). Thus, the projection of x onto the regression line (which is the conditional expected value of x, $\hat{x}$) is uncorrelated with z. Panels b and c show that, while x is on average lower for $z = 0$ than for $z = 1$, $\hat{x}$ has a very similar distribution for the two values of z. Finally, Figure 7.4(d) shows that the second-stage regression is very close to the generating model evaluated at the mean value of z, which is the relationship that we are actually trying to recover. Contrast this with the simple regression of y on x shown in Figure 7.2.

Getting correct standard errors of the parameter estimate requires some more work, so in practice we use software that does this automatically. Traditionally economists use Stata, which has lots of tools for instrumental variable regression. In R, the two functions that have been developed are `ivreg()` in the `AER` library (Kleiber and Zeileis 2008) and `tsls()` in the `sem` library (Fox et al. 2013). The former (which I use in this chapter) has a slightly simpler syntax; for basic regressions it works just like `lm()` except that the formula includes a vertical bar followed by the instrumental variables; e.g., `ivreg(y ~ x | u)`. The fitted model can be viewed with `summary()`, generating output very similar to `lm()`. Parameter estimates and standard errors for the full OLS regression, the short-form regression, and the 2SLS regression are shown in Table 7.1. Although the standard error on the 2SLS slope estimate is large, the estimate itself is clearly better than the short-form regression (the error in the intercept might be improved by **centering** the data — see Chapter 10).

Coefficient	Generating model	OLS y~x+z	OLS y~x	2SLS y~x|u
Intercept	4	3.99 (0.29)	3.57 (0.61)	6.24 (0.82)
x	-2	-1.87 (0.61)	3.93 (1.13)	-1.35 (1.55)
z	5	4.72 (0.25)	–	–

Table 1: Table 7.1: Results of fitting the data in Figure 7.2 using OLS with the generating model and without the omitted variable, and using two-stage least squares with an instrumental variable. The "Generating model" column shows the values of the coefficients used to generate the simulated data.

7.3.1 Simultaneous equation models

Bidirectional causality is best solved through a system of simultaneous equation models. For example, in examining the effects of disease on economic outcomes, Bonds et al. (2012) write down two equations, one for disease burden as a function of income and some other variables, and the other for income as a function of disease burden and some other variables. As discussed above, there is endogeneity between disease burden and income, so these equations cannot be estimated via ordinary least squares. Instead, instruments are
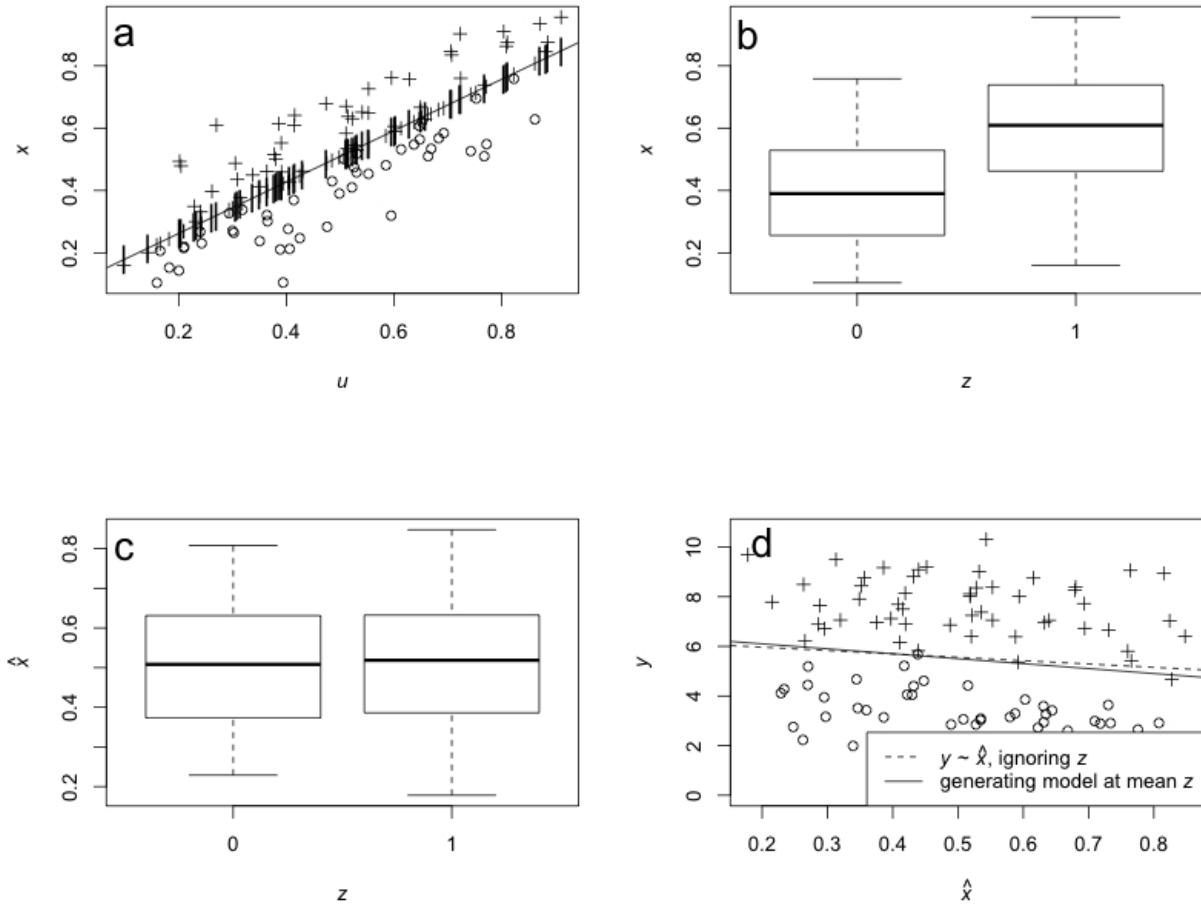
Figure 4: *Figure 7.4:* Instrumental variable regression of the data in Figure 7.2, using an instrument u that is positively correlated with x but uncorrelated with z. (a) Relationship between x and u, showing the data (circles and plusses) and the fitted values ($hatx$; vertical hashes on the diagonal). Plusses: $z = 1$; circles: $z = 0$. (b) On average, x is larger when $z = 1$ than when $z = 0$. (c) In contrast, $\hat{x}$ is uncorrelated with z. (c) The regression of y on $\hat{x}$ is very similar to the generating model evaluated at the mean value of z.

10

identified for both income and disease burden; and in the equation for disease burden, for example, the income variable is replaced with its predicted value from the first-stage regression. The same is done for the other equation.

This model can be estimated using two-stage least squares, as for the single-equation model. However, another technique, called three-stage least squares, is more efficient; but it requires additional restrictions on the instrumental variables (notably, that they be uncorrelated with the residuals from *all* of the equations, not just the equation to which they are being applied).

The easiest way to fit these types of models in R is with the `systemfit` library. The documentation for this package (Henningsen & Hamann 2007) has a good overview of the various statistical approaches to fitting these models.

7.4 Life history tradeoffs in Florida Scrub-Jays

As a concrete example of the life-history tradeoff problem, let's look at some data from the population of Florida Scrub-Jays (*Aphelocoma coerulescens*) that has been studied at Archbold Biological Station since the early 1970s (Woolfenden & Fitzpatrick 1984). The focus here is on the tradeoff between current-year reproductive success and survival over the succeeding winter. For reasons that will become clear in a bit, I focus on reproductive males.

The file `JayData.csv` has the following variables:

- `fledge_number`: the number of offspring that a breeding pair fledged in a given year. This is our measure of reproductive success
- `dad_survival`: a zero/one variable indicating whether the male of the breeding pair survived to the subsequent breeding season. This is our measure of subsequent fitness.
- `hatching_success`: a zero/one variable indicating whether the breeding pair hatched any chicks in the year. This will be our instrumental variable.

First lets look at a simple OLS regression.

```
> jays <- read.csv("JayData.csv")
> jay_lm <- lm(dad_survival ~ fledge_number, data=jays)
> summary(jay_lm)

Call:
lm(formula = dad_survival ~ fledge_number, data = jays)

Residuals:
    Min      1Q  Median      3Q     Max
-0.7881  0.2232  0.2288  0.2458  0.2458

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.754193   0.019067  39.555   <2e-16 ***
fledge_number 0.005659   0.008024   0.705    0.481
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.4246 on 1204 degrees of freedom
Multiple R-squared: 0.0004128,  Adjusted R-squared: -0.0004174
F-statistic: 0.4973 on 1 and 1204 DF,  p-value: 0.4808
```

Thus, subsequent survival is positively related to fledgling number, although the coefficient is not significantly different from zero. Given that the response variable is binary, one might reasonably argue that logistic regression is more appropriate for these data; however, the consistency guarantee of IV regression only applies to least squares models. Just to check that the least squares assumption isn't too egregious, let's look at a logistic regression; the result is qualitatively similar (the magnitudes of the coefficients are different, because of the transformation of the response variable; but the *P*-value for the effect of `fledge_number` is identical):

```
> jay_glm <- glm(dad_survival ~ fledge_number, data=jays, family=binomial)
> summary(jay_glm)

Call:
glm(formula = dad_survival ~ fledge_number, family = binomial,
    data = jays)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7599   0.7109   0.7208   0.7512   0.7512

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.12077    0.10465  10.709   <2e-16 ***
fledge_number  0.03149    0.04463   0.705    0.481
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1316.5  on 1205  degrees of freedom
Residual deviance: 1316.0  on 1204  degrees of freedom
AIC: 1320

Number of Fisher Scoring iterations: 4
```

So does this mean that there is no tradeoff? Well, as described aboce, there might be differences in individual qualitiy that affect both fecundity and survival, but in this dataset we have no measures of quality that we could include in the regression. So now let's look for in instrumental variable: a quantity that is correlated with fledgling number, but has no direct causal effect on survival, and is not correlated with individual quality.

The Florida Scrub-Jay is a social breeder, and the presence of helpers is known to increase reproductive success (Woolfenden & Fitzpatrick 1984). It is at least plausible that helpers have little affect on parental survival. However, helpers are the offspring of the breeders from previous years of successfull reproduction, so that breeders that have higher than average fledgling numbers, because of being of higher quality, are more likely to have helpers. Thus the presence of helpers is probably correlated with breeder quality, disqualifying them as an instrument.

Another candidate variable is nest failure, that is, the loss of all eggs in the nest prior to hatching, whether because of predation or because of the nest falling to the ground. Certainly we would expect this variable to be correlated with fledgeling number, as the latter will always be zero when no eggs hatch (this species will sometimes renest; here I use a "hatching success" variable that is zero only if all nests fail to hatch any chicks). Female breeders will have expended some energy laying these eggs, so their reproducitive investment is not zero; but male breeders whose nests fail to hatch have expended almost no reproductive energy. If hatching failure is effectively random across male breeders, then by comparing the survival of those who hatched chicks and those who did not should reveal the average cost of raising an average-sized brood (conditioned on hatching).

The question is, is hatching success correlated with male quality? Probably not with energetic or physiological quality; but predator defense ability is a form of quality that might affect both fledgling number and individual survival. In addition, helpers may help defend the nest from predators, and as mentioned above, the presence of helpers is probably associated with breeder quality. Thus, hatching success is at best an imperfect instrument; but it is probably less correlated with the omitted quality variable than raw fledgling number. Thus we will try using it as an intstrumental variable in a two-stage least squares.

```
> library(AER)
> jay_2sls <- ivreg(dad_survival ~ fledge_number | hatching_success, data=jays)
> summary(jay_2sls)

Call:
ivreg(formula = dad_survival ~ fledge_number | hatching_success,
    data = jays)

Residuals:
    Min      1Q  Median      3Q     Max
-0.8209  0.1791  0.2100  0.2719  0.4265

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.82090    0.03699  22.191   <2e-16 ***
fledge_number -0.03092    0.01913  -1.617    0.106
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.4282 on 1204 degrees of freedom
Multiple R-Squared: -0.01684,   Adjusted R-squared: -0.01769
Wald test: 2.614 on 1 and 1204 DF,  p-value: 0.1062
```

The result is suggestive of a tradeoff: every extra fledgling reduces male survival probablitiy by 0.031, or 3.1 percentage points. The standard errors are relatively large, and under ecological conventions for α levels we cannot reject the null hypothesis of no effect. Nevertheless, we have relatively high confidence that the effect is not negative, and theory says that this estimate is more reliable than the one from the OLS regression.

7.5 Other issues with instrumental variable regression

The mathematical proofs that IV regression gives consistent parameter estimates depends both on the linearity of the response and the use of least squares, so IV cannot be used with generalized linear regression or nonlinear regression. This is why I used least squares in the scrub-jay analysis, even though the response is a dichotomous variable. With some care, this is not as much of a shortcoming as it might seem. For example, as long as the fitted probabilities are not too close to zero or one, the assumption of linearity on an untransformed scale is not too bad (and for other types of GLM, an appropriate tranformation can often be applied to the response variable to linearize the relationship). The only effect of applying least squares when the theoretical residuals are not normally distributed is to genearate biased standard errors — the parameter estimates are consistent and unbiased — and this effect gets small as sample size gets large.

Although the IV estimates are consistent (as long as the instrument truly meets the criteria outlined above), they can still be biased with a finite sample size. In general, for a given sample size, this bias is lower the better the instrument is at predicting the endogenous variable. A commonly cited rule of thumb is that the F statistic for the first-stage regression should be at least 10 for the instrument to be useful. If the F statistic is lower than that, then it is called a "weak instrument," and the estimates from the IV regression are likely to be no more reliable than the inconsistent estimates from the OLS regression. If there are multiple

instruments, and some of them are very weak, then retaining the the IV regression may tend to increase the bias. In that case, it may be beneficial to prune variables from the first stage regression in much the same way one would do for model selection in ordinary multiple regression (with the caveat that the exogenous variables must be left in).

We run this test simply by explicitly running the first stage regression using OLS. Here is the scrub jay example:

```
> jay_1st <- lm(fledge_number ~ hatching_success, data=jays)
> summary(jay_1st)

Call:
lm(formula = fledge_number ~ hatching_success, data = jays)

Residuals:
    Min      1Q  Median      3Q     Max
-2.0513 -1.0513  0.0000  0.9487  5.9487

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)      -8.184e-15  1.194e-01     0.0        1
hatching_success  2.051e+00  1.266e-01    16.2   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.382 on 1204 degrees of freedom
Multiple R-squared: 0.179,  Adjusted R-squared: 0.1784
F-statistic: 262.6 on 1 and 1204 DF,  p-value: < 2.2e-16
```

We see that $F_{1,1204} = 262$, so bias will not be an issue here.

Starting with version 1.2, the AER package has a helpful diagnostics argument to the summary() function for IV regression objects:

```
> if (packageVersion("AER") >= 1.2) summary(jay_2sls, diagnostics=TRUE)

Call:
ivreg(formula = dad_survival ~ fledge_number | hatching_success,
    data = jays)

Residuals:
    Min      1Q  Median      3Q     Max
-0.8209  0.1791  0.2100  0.2719  0.4265

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.82090    0.03699  22.191   <2e-16 ***
fledge_number -0.03092    0.01913  -1.617    0.106

Diagnostic tests:
                 df1  df2 statistic p-value
Weak instruments   1 1204   262.577  <2e-16 ***
Wu-Hausman         1 1203     4.546  0.0332 *
Sargan             0   NA        NA      NA
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.4282 on 1204 degrees of freedom
Multiple R-Squared: -0.01684,   Adjusted R-squared: -0.01769
Wald test: 2.614 on 1 and 1204 DF,  p-value: 0.1062
```

The thing to look at here is the statistic reported for the "Weak instruments" test — it is F from the first stage regression that we calculated above (there are a variety of ways of calculating P-values for the weak instruments test; the precise meaning of the P-value reported here is not documented). The "Wu-Hausman" and "Sargan" tests are discussed below.

The examples so far have had a single predictor variable, which was endogenous. More generally, we will want to perform regressions with multiple predictor variables, some of which may be truly endogenous. To do instrumental variable regression in this case, we need at least as many instruments as there are endogenous variables; and the exogenous variables need to be included in both the first stage and second stage regressions. Thus, for example, if x1 and x2 were exogenous, x3 and x4 were endogenous, and u1, u2, and u3 were each instruments for one or the other (or both) endogenous variables, then the call to `ivreg` would be `ivreg(y ~ x1 + x2 + x3 + x4 | x1 + x2 + u1 + u2 + u3)`. My economics colleagues tell me that it is relatively rare to instrument more than one endogenous variable in a single equation, however, as it is quite difficult to interpret the result unless the problem is set up very carefully (e.g., Angrist 2010).

It is often the case that there are more instruments than endogenous variables. This is fine; subject to the caveats about very weak instruments above, to the extent that multiple instruments improves the explanatory power of the first stage regression, they will improve the final model. Having multiple instruments also allows one to conduct "overidentification" tests, the most common of which is the Sargan test reported by the `AER` package. Under the assumption that at least one of the instruments is exogenous, then the Sargan test tests the null hypothesis that all of the other instruments are also exogenous. A small P-value from this test suggests that at least one of the instruments is endogenous and fails the exclusion condition; in contrast, a large P-value can be evidence of valid evidence. However, without *a priori* confidence that at least one of the instruments is valid, the outcome of this test is meaningless.

7.6 Deciding to use instrumental variable regression

Although the computational mechanics of running an instrumental variable regression are quite simple, the identification of observed variables that meet the assumptions required to serve as instruments is often quite challenging. Thus IV regression should not be undertaken lightly. Unfortunately, there is no simple diagnostic test of your original data that will tell you whether your predictor variables are endogenous. Endogeneity means that the predictor variable is correlated with the actual residuals, but the latter are not observable; OLS regression, by its very design, produces estimated residuals that have zero linear correlation with the predictor variables. As I suggested above, one might imagine that a single omitted variable with a single predictor variable might produce heteroskedasticity in the skew of the estimated residuals; but I know of no statistical theory or practice to support that conjecture, and even if it applies, it will not help with other sources of endogeneity.

The only solution is to think hard and critically about potential relationships among your variables, just as we do when designing field experiments (are the quadrats truly randomly distributed with respect to the statistical population that I want to draw inferences about? How will the raptor perches created by my exclosures affect the ecological processes I am interested in?). Here are some questions that you can ask about your data:

- What is the likely magnitude of measurement error in my predictor variables? Can I estimate it, and can I apply one of the other techniques for correcting measurement error bias?

- What are all of the variables processes that might causally affect my response variable? Of the ones whose effects are not guaranteed to be small, how might they be correlated with the predictor variables that I actually want to draw inferences about?
- Are there any feedback loops in my system? If so, are there plausible pathways whereby my response variable might directly or indirectly affect my predictor variable?

If these questions lead to the conclusion that endogeneity is likely in your model, then you should try to identify one or more instrumental variables.

Once you have identified some instrumental variables, and you are confident that they are good instruments, then you can use them to test whether your predictor variable really is endogenous. This is important to do because if the predictor variable is actually exogenous, then the OLS regression is already consistent and the IV regression is both less efficient (confidence intervals will be wider) and more prone to bias than the OLS regression.

The standard test for endogeneity is the Wu-Hausman test. The basic idea of this test is that if the predictor variable (x) is exogenous, then the residuals of the first stage regression describing the relationship between the instruments and the predictor variable ($x \sim u$) will be uncorrelated with the response variable (y). In practice the test is performed by taking the estimated residuals from the first stage regression and adding them as an additional variable in structural equation; then using OLS regression to test the null hypothesis that the coefficient of these residuals is zero. Failing to reject this null hypothesis leads to the conclusion that the predictor variable is exogenous; rejection leads to the conclusion that the predictor variable is endogenous, and IV regression is required.

Here is the procedure for the scrub jay example:

```
> jay_1st <- lm(fledge_number ~ hatching_success, data=jays)
> jay_wh <- lm (dad_survival ~ fledge_number + resid(jay_1st), data=jays)
> summary(jay_wh)

Call:
lm(formula = dad_survival ~ fledge_number + resid(jay_1st), data = jays)

Residuals:
    Min      1Q  Median      3Q     Max
-0.8209  0.1791  0.2296  0.2432  0.2705

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.82090    0.03662  22.415   <2e-16 ***
fledge_number -0.03092    0.01894  -1.633   0.1027
resid(jay_1st) 0.04456    0.02090   2.132   0.0332 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.4239 on 1203 degrees of freedom
Multiple R-squared: 0.004176,   Adjusted R-squared: 0.00252
F-statistic: 2.522 on 2 and 1203 DF,  p-value: 0.0807
```

The P-value for the first stage residuals is 0.033, so we reject the null hypothesis of exogeneity, and conclude that the IV regression is the appropriate analysis for this model. Notice that this P-value is reported by `diagnostics=TRUE` argument to the `summary()` function for IV regression objects (see section 7.5).

Of course, if the instruments are not appropriate — they are poor predictors of the predictor variable (which can be seen in the first stage regression) or are not themselves exogenous (which can only be evaluated through reasoning) — then the results of the Wu-Hausman test are meaningless.

7.7 Choosing instrumental variables

How does one find a good instrument? It is relatively easy to determine whether a candidate instrument is strongly correlated with the endogenous variable — that is a simple statistical test. However, the second criterion, that the instrument affects the response variable *only* through its effect on the endogenous variable, requires careful thought. The case of endogeneity arising from measurement error is probably the easiest, as it is relatively easy to find an instrument that is uncorrelated with the measurement error process. However, even here it is critical to ensure that there are no other pathways by which the instrumental variable can influence the response variable.

One fairly straightforward situation is when addressing endogeneity in an experiment, where the experimental treatment is not the predictor variable (e.g., the hypothetical phytoplankton-*Daphnia* experiment, where phytoplankton density was the predictor variable but it was manipulated indirectly using variable nutrient input). Here, the experimental treatment itself is a clear instrumental variable: it is correlated with the predictor variable (if it was not, then the experiment would have failed), and has no direct effect on the response variable (if it does, then the whole rationale for the experiment breaks down). Basically, you are predicting the response variable using the *expected* value of the predictor variable, conditioned on the experimental treatment, rather than the actual value of the predictor variable. This removes the potential correlation between the responses of the predictor variable and the dependent variable to random variation among experimental units.

In a time series context, *lagged* variables can often fit the bill. This is used by Creel & Creel (2009) in estimating the population dynamics of Rocky Mountain elk. The authors fit a density-dependent Gompertz model,

$$r_t = \log(N_{t+1}/N_t) = a + b \log(N_t) + \beta E_t + \epsilon_t,$$

where E is a vector of environmental variables such as snow depth and wolf density. The problem is that population abundance, N_t, is measured with substantial uncertainty, and this measurement error will tend to bias the density dependence parameter (b) towards zero. The authors addressed this with two techniques, instrumental variable regression and state-space modeling, and found that they gave similar results. As instruments they used lagged values of the environmental variables, E_{t-1} and E_{t-2}; the authors do not explicitly state their reasoning, but it would typically be that such variables may affect the "true" value of N_t (through their effects on past population dynamics) but have no direct effect on either the current population growth rate or the measurement error. As is common practice in economics when instruments are generic rather than carefully crafted to emulate an experiment, squares and interactions of the lagged environmental variables are included as well. The authors did not report the F-statistic of the first-stage regression, but noted that the $R^2 = 0.20$, which they consider to be "large considering the overall noise in the data."

The only other example of instrumental variables in the ecological literature at the time of writing draws heavily on economic reasoning. Hanley et al. (2008) evaluate the effects of grazing on plant biodiversity over the past 400 years in Scotland. Although historical plant diversity could be reconstructed from pollen records, there were not data on stocking rates over the entire study period. The authors used meat price as a proxy variable, arguing that no single farmer influenced price and so would increase stocking rate when prices were high without endogenously affecting price. However, there are times when price is high because of a regional collapse of stock, perhaps due to an unusually harsh winter or a disease outbreak. Thus, the authors sought instrumental variables that would indicate an unusual spike in demand or trough in supply: "the English population (a measure of market demand); the presence or absence of garrisons in a particular region (which increases local demand); the passing of the Act of Union between England and Scotland (this reduced trade barriers); the advent of refrigerated transport from the New World (this reduced demand for UK meat); and grain prices, as a substitute for meat in consumption." These were used as instruments for meat price in the first stage regression, such that the fitted value of price from the first stage model can be treated as an exogenous proxy for stocking rates in the second stage. Note that dealing with this endogeneity between supply, demand, and price is the bread and butter of econometrics.

An important use of instrumental variables in the social sciences is in program evaluation: e.g., determining whether a public health program targeted at poor people results in improved health outcomes. The challenge comes from potential selection bias: the service is more likely to be used by people in especially poor health,

who may still be sicker than average after treatment. The solution is to use *program eligibility* to instrument program participation, essentially comparing outcomes for those who were eligible, regardless of participation, with those who were not. This works as long as there is some source of arbitrariness in the eligibility rule. If the rule is issued based on the same factors that may affect your outcome of interest, then it is not a good instrument. Of course, such analyses need to control for a variety of other variables as well. Similar analyses have been done on the effects of higher education on subsequent earning, using access to education as an instrument for actual enrollment to address the endogeneity arising from the fact that individuals who are ambitious enough to go to college may be predisposed to success in the workplace. Similar approaches may be useful to evaluate the ecological effects of various human activities. For example, the effects of marine protected areas on biodiversity and species abundance are usually assessed by comparing measurements inside an MPA with a nearby site that is not protected. However, within a planning area, the MPAs are not sited at random, and may be consistently higher or lower in recovery potential than sites that are not protected. One can imagine instrumenting protection status of a site with "eligibility to be protected" — i.e., is the site inside the planning area that was being considered when setting up the MPAs? Of course, this requires that there be sites outside the planning area that are broadly ecologically comparable to those in the planning area, but in many cases that may be possible.

The biggest challenging in determining whether a variable is a suitable instrument is verifying the exogeneity criterion. Just as in the initial assessment of potential sources of endogeneity, this requires a critical analysis of potential causal pathways, based on a scientific understanding of the system. For example, in the elk model, one can imagine that the previous year's precipitation could affect vegetation structure and greenness in the current year, which could, in turn, affect the detectability of elk (and hence the measurement error residual). The plausibility of such a scenario would have to be evaluated using prior knowledge about the effects of rainfall on vegetation and about the effects of vegetation on detestability. An ideal analysis would examine such potential critiques with as much quantitative and qualitative data as is available.

7.8 Conclusion

In summary, instrumental variable regression may provide a way to develop robust causal interpretations of patterns in observational ecological datasets. The challenge in this technique is not technical, but in the conceptual identification of suitably independent instruments. Until we ecologists become more practiced at thinking through all of the ways that putative instruments might have hidden correlations with the factors causing endogeneity, it may be best to discuss any such analyses with economist colleagues (as was done, for example, by Hanley et al. 2008 and Creel & Creel 2009). For those who want to delve more deeply into the subject, some useful econometrics textbooks with extensive treatments of instrumental variables include Angrist and Pischke (2009) and Wooldridge (2010).

A general checklist for doing instrumental variable regression is as follows (relevant chapter sections are in parentheses):

1. Determine plausible or known sources of endogeneity in your predictor variables, considering measurement error, omitted variables, and bidirectional causality (7.2.1)
2. Identify potential instrumental variables (7.7)
3. Run the first stage regression to determine if the instruments are strong enough to be useful ($F > 10$) (7.5)
4. If you have multiple predictor variables, run some identification checks to see if some of them are redundant or so weak that they can be dropped (7.5)
5. Run the Wu-Hausman test of endogeneity (7.6)
6. If the instruments are sufficiently strong, and the Wu-Hausman test confirms endogeneity, then run the IV regression. Otherwise, stick with your original OLS regression (7.3).

In this chapter, I have focused on consistent estimates of regression parameters. However, endogeneity also renders P-values meaningless, precluding hypothesis tests and qualitative assessments of causality.

Instrumental variable regression solves this problem: although the power can be substantially reduced, the P-value is accurate and can be used in statistical hypothesis tests. Furthermore, by effectively randomizing with respect to the potential sources of endogeneity, a well-constructed instrumental variable provides as robust a demonstration of causality as does a well-designed experiment.

Acknowledgements

References

Angrist, J. D. 2010. Multiple endogenous variables—now what?! Mostly Harmless Economics Blog. Online resource http://www.mostlyharmlesseconometrics.com/2010/02/multiple-endogenous-variables-what-now/ [accessed 10 January 2014].

Angrist, J. D., & Pischke, J.-S. (2009). Mostly harmless econometrics. Princeton, NJ: Princeton University Press.

Armsworth, P. R., Gaston, K. J., Hanley, N. D., & Ruffell, R. J. (2009). Contrasting approaches to statistical regression in ecology and economics. Journal of Applied Ecology, 46(2), 265-268. doi: 10.1111/j.1365-2664.2009.01628.x

Bonds, M. H., Dobson, A. P., & Keenan, D. C. (2012). Disease ecology, biodiversity, and the latitudinal gradient in income. Plos Biology, 10(12), e1001456. doi: 10.1371/journal.pbio.1001456

Buonaccorsi, J. P. (2010). Measurement error: Models, methods, and applications. Boca Raton, FL: Chapman and Hall/CRC.

Creel, S., & Creel, M. (2009). Density dependence and climate effects in Rocky Mountain elk: an application of regression with instrumental variables for population time series with sampling error. Journal of Animal Ecology, 78(6), 1291-1297. doi: 10.1111/j.1365-2656.2009.01581.x

de Jong, G., & van Noordwijk, A. J. (1992). Acquisition and allocation of resources: genetic (co)variances, selection, and life histories. American Naturalist, 139(4), 749-770. doi: 10.1086/285356

Fox, J., Nie, Z., & Byrnes, J. (2013). sem: Structural Equation Models (Version 3.1-3). Retrieved from http://CRAN.R-project.org/package=sem

Freckleton, R. P. (2011). Dealing with collinearity in behavioural and ecological data: model averaging and the problems of measurement error. Behavioral Ecology and Sociobiology, 65(1), 91-101. doi: 10.1007/s00265-010-1045-6

Greene, W. H. (2008). Econometric analysis (6th ed.). Upper Saddle River, N.J.: Pearson/Prentice Hall.

Hanley, N., Davies, A., Angelopoulos, K., Hamilton, A., Ross, A., Tinch, D., & Watson, F. (2008). Economic determinants of biodiversity change over a 400-year period in the Scottish uplands. Journal of Applied Ecology, 45(6), 1557-1565. doi: 10.1111/j.1365-2664.2008.01570.x

Henningsen, A., & Hamann, J. D. (2007). systemfit: A package for estimating systems of simultaneous equations in R. Journal of Statistical Software, 23(4), 4.

Kleiber, C., & Zeileis, A. (2008). Applied econometrics with R. New York: Springer Verlag.

Morgan, M. S. (1990). The history of econometric ideas. Cambridge ; New York: Cambridge University Press.

Solow, A. R. (1998). On fitting a population model in the presence of observation error. Ecology, 79(4), 1463-1466. doi: 10.1890/0012-9658(1998)079[1463:ofapmi]2.0.co;2

Spiller, D. A., & Schoener, T. W. (1995). Long-term variation in the effect of lizards on spider density is linked to rainfall. Oecologia, 103(2), 133-139. doi: 10.1007/bf00329072

Spiller, D. A., & Schoener, T. W. (1998). Lizards reduce spider species richness by excluding rare species. Ecology, 79(2), 503-516. doi: 10.1890/0012-9658(1998)079[0503:lrssrb]2.0.co;2

Stock, J. H., & Trebbi, F. (2003). Who invented instrumental variable regression? Journal of Economic Perspectives, 17(3), 177-194. doi: 10.1257/089533003769204416

Underwood, A. J. (1997). Experiments in Ecology: Their Logical Design and Interpretation Using Analysis of Variance. Cambridge, UK: Cambridge University Press.

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data (2nd ed.). Cambridge, Mass.: MIT Press.

Woolfenden, G. E., & Fitzpatrick, J. W. (1984). The Florida scrub jay: demography of a cooperative-breeding bird. Princeton, N.J.: Princeton University Press.