

The Other Transformation in Econometric Practice: Robust Tools for Inference

Author(s): James H. Stock

Source: *The Journal of Economic Perspectives*, Vol. 24, No. 2 (Spring 2010), pp. 83-94

Published by: American Economic Association

Stable URL: <http://www.jstor.org/stable/25703501>

Accessed: 11-05-2016 18:46 UTC

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at

<http://about.jstor.org/terms>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



American Economic Association is collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Economic Perspectives*

The Other Transformation in Econometric Practice: Robust Tools for Inference

James H. Stock

In the early 1980s, a trio of prominent papers laid out a scathing critique of contemporary econometric practice. In “Macroeconomics and Reality,” Christopher Sims (1980) argued that the “incredible” exclusion restrictions used to estimate standard large macroeconomic models undercut the reliability of policy advice based on those models; he endorsed instead using as few identifying assumptions as possible and proposed vector autoregressions as an alternative modeling strategy. In his inaugural lecture at the London School of Economics, “Econometrics: Alchemy or Science?” David Hendry (1980) recited a lengthy list of pitfalls of regression studies, most of which can be interpreted in current terminology as threats to the identification of the effect of interest; he then sketched how research can be conducted using a minimum of identifying restrictions and ended up with what can be seen in retrospect as an error-correction/cointegration model of money demand.¹ In “Let’s Take the Con Out of Econometrics,” Edward Leamer (1983) attacked both the “whimsical” nature of the assumptions used to justify inferences in econometric regression studies using observational data and the fragility of results to arbitrary decisions about choice of control variables; he then proposed extreme bounds analysis as a tool for quantifying the sensitivity of regression

¹ In an error-correction model, the change in a time-series variable depends on the lagged gap between the levels of that variable and one or more other variables; this gap is the “error correction” term. As an aside, around the time of Hendry’s lecture Clive Granger set out to prove such models were internally inconsistent but in fact proved the opposite, and his resulting characterization of such processes, which he called “co-integrated,” ultimately won him the 2003 Nobel Prize in Economics (Granger, 2003).

■ *James H. Stock is Harold Hitchings Burbank Professor of Political Economy, Harvard University, Cambridge, Massachusetts. He is also a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts.*

doi=10.1257/jep.24.2.83

estimates. These three articles targeted different audiences and proposed quite different techniques for solving their perceived deficiencies in current practice. But they shared the same core message: far more attention needed to be paid to identification of the causal effect of interest, and econometric inference should not hinge on subsidiary modeling assumptions.

These two objectives—first, credible identification of key causal effects or parameters, and second, statistical inference that is robust to subsidiary modeling assumptions—have guided much of applied and theoretical econometric research since these three papers. In hindsight, this trio’s critiques were on target. Moreover, while there have inevitably been some false turns and dead ends, the applied and theoretical econometric research flowing in part from those papers has produced important advances that serve to make empirical work today far more credible than it was three decades ago.

The combined effect of this research has been to transform econometric practice and teaching—in my view, very much for the better. A modern undergraduate course in econometrics looks and feels very different than it did 20 or even ten years ago. In the 1980s, the standard undergraduate curriculum addressed estimation and inference in models in which all regressors were treated symmetrically. In contrast, the classroom today focuses more on the estimation of specific objects, mainly specific causal effects, and less on the estimation of “models.” For example, last spring I supervised an undergraduate senior thesis with the goal of estimating the effect of steroids on home runs in Major League Baseball (a tricky task, since we don’t have player-level medical test data on steroid use); 20 years ago, the goal might have been to develop a model of home runs. The difference in emphasis—a specific causal effect instead of “estimating a model”—reflects a helpful narrowing of scope that makes one focus on the key sources of identification.

Angrist and Pischke’s article (this issue) highlights one aspect of the first of these two research strands listed above—specifically, the rise of experiments and quasi-experiments as credible sources of identification in microeconomic studies, which they usefully term “design-based research.” But in so doing, they miss an important part of the story: the second research strand aimed at developing tools for inference that are robust to subsidiary modeling assumptions. My first aim in these remarks therefore is to highlight some key developments in the second research strand. I then turn to Angrist and Pischke’s call for adopting experiments and quasi-experiments in macroeconometrics; while sympathetic, I suspect the scope for such studies is limited. I conclude with some observations on the current debate about whether experimental methods have gone too far in abandoning economic theory.

Credible Inference

The past three decades have seen significant changes in the tools of econometrics, many motivated by a desire to minimize the effect of “whimsical” assumptions on inference about the object of interest. By “whimsical” I mean arbitrary assumptions

that are subsidiary to the empirical purpose at hand, but which affect inference about the causal effect of interest.² The new tools provide reliable inference without implausible subsidiary assumptions. I illustrate this by four examples: robust standard errors, methods for inference with weak instruments, handling of control variables, and nonparametric and semiparametric regression.

Robust Standard Errors

The standard errors conventionally provided in an ordinary least squares regression analysis are based on the assumption that the error term in the regression is homoskedastic, that is, has a variance that does not depend on the regressors and is the same for all observations. When this assumption is violated, heteroskedasticity arises, in which case the estimated regression coefficients are unaffected but the standard errors and statistical significance are unreliable.

The 1970s procedure for handling potential heteroskedasticity was either to ignore it or to test for it, to model the variance as a function of the regressors, and then to use weighted least squares. While in theory weighted least squares can yield more statistically efficient estimators, modeling heteroskedasticity in a multiple regression context is difficult, and statistical inference about the effect of interest becomes hostage to the required subsidiary modeling assumptions. White's (1980) important paper showed how to get valid standard errors whether there is heteroskedasticity or not, without modeling the heteroskedasticity. This paper had a tremendous impact on econometric practice: today, the use of heteroskedasticity-robust standard errors is standard, and one rarely sees weighted least squares used to correct for heteroskedasticity.

The widespread adoption of heteroskedasticity-robust standard errors has also alleviated one of the more awkward moments in introductory undergraduate econometrics: when the teacher defines "homoskedasticity"; explains that we will assume it for now, though it isn't really true, of course; and promises to deal with this later in the semester. In the circa-1980 textbooks, the assumption of homoskedasticity was "relaxed" in a later chapter using weighted least squares (for examples of three leading books at the time, see Pindyck and Rubinfeld, 1981; Gujarati, 1978; Wonnacott and Wonnacott, 1979). Today, there is no need to introduce the assumption of homoskedasticity in the first place: one can simply follow common practice and use heteroskedasticity-robust standard errors from the outset.

Work on robust inference is ongoing, and there have been recent important developments in robust standard errors for panel data regression. In panel data, the errors for a given entity (individual) might be serially correlated, and if so, then conventional ordinary least squares standard errors are unreliable. Bertrand, Duflo, and Mullainathan (2004) brought this problem to the attention of the

² Here is Leamer's (1983, pp. 37–38) definition of whimsical: "Sometimes I take the error terms to be correlated, sometimes uncorrelated; sometimes normal and sometimes nonnormal; sometimes I include observations from the decade of the fifties, sometimes I exclude them; sometimes the equation is linear and sometimes nonlinear; sometimes I control for variable z , sometimes I don't."

applied community by demonstrating, in an empirically motivated Monte Carlo study, that conventional ordinary least squares standard errors for fixed effects regression can substantially understate the sampling uncertainty. In panel data, if the errors are uncorrelated across entities, then inference that is robust to general heteroskedasticity and to serial correlation within an entity can be conducted using “clustered” standard errors, where in this case the entity is the “cluster.” It has been known for some time that clustered standard errors are consistent under potential serial correlation and heteroskedasticity if the number of clusters (entities) is large and the number of observations per cluster (time periods) is small (Kiefer, 1980; Arrelano, 1987, 2003). But recent work by Hansen (2007) has shown that, surprisingly, valid inference remains possible with general serial correlation even if the number of time periods is large and the number of entities small. Even though the variance is imprecisely estimated in this case, t - and F -statistics have simple distributions, so hypothesis tests and confidence intervals based on clustered standard errors remain valid as long as one uses the right critical values (t - and F -tables with additional degrees-of-freedom adjustments).³

Inference with Possibly Weak Instruments

One hallmark of the new design-based research advocated by Angrist and Pischke is close scrutiny of proposed instrumental variables so that they do in fact provide credible identification. The question of whether instruments plausibly capture truly exogenous variation—variation unrelated to the error in the regression of interest—is at center stage. One unintended consequence of this focus on credibly exogenous variation is that, in many applications, these instrumental variables are found to have a weak correlation with the included endogenous regressors for which they are instruments, conditional on the control variables. This weak correlation raises technical problems for conducting statistical inference; in particular, the usual textbook asymptotic normal and chi-squared distributions of instrumental variable regression statistics can in this case provide poor approximations to sampling distributions, even if the sample size is large.

Fortunately, econometric theorists have been working hard on this so-called weak instruments problem and have developed a suite of tools for inference in instrumental variables regression (and generalized method of moments estimation) when the instruments are possibly weak. Some of these tools are becoming sufficiently well-accepted that they are commonplace in empirical work published in top journals and are starting to appear in undergraduate textbooks. The simplest of these tools is to look at the first-stage F -statistic—that is, the F -statistic testing

³ Hansen’s (2007) result is akin to the notion that in finite samples with normally distributed data, valid hypothesis tests and confidence intervals can be constructed using the t -distribution, even though the error variance is not consistently estimated. Hansen’s result applies to a balanced panel with independence across entities and allows for general serial correlation and heteroskedasticity that has the same form across entities. These assumptions are appropriate if the entities are drawn using simple random sampling. Ibragimov and Müller (forthcoming) provide a different approach to robust inference in panel data that relaxes the assumption of homogeneity across entities.

the hypothesis that the coefficients on the instruments are zero in the first-stage regression of two-stage least squares. If this F -statistic is large—a common rule of thumb is $F > 10$ —then one can treat the instruments as sufficiently strong that the usual two-stage least squares output can be used.

If the first-stage F -statistic is small, however, then two-stage least squares can be badly biased and the accompanying confidence intervals can be quite misleading (typically, too short and centered in the wrong place—a bad combination!). The econometric literature on what to do if you have weak instruments is large and there is not room here to do it justice, so I mention only two developments. First, a remarkable result in this recent literature on weak instruments is that, if you have a single included endogenous regressor, Moreira's (2003) conditional likelihood ratio statistic effectively produces valid and fully efficient confidence intervals and hypothesis tests regardless of whether instruments are weak, strong, or even irrelevant. Second, one estimation method that performs better than two-stage least squares when instruments are weak is limited information maximum likelihood (LIML), although LIML can produce extreme estimates. Commands for inference in instrumental variable regression with possibly weak instruments are increasingly available in statistical software packages; for example, first-stage F -statistics, conditional likelihood ratio statistic confidence intervals, and the LIML estimator are readily computed in STATA.

Control Variables

Another awkward moment in undergraduate instruction comes when a clever undergraduate asks why we are interpreting the coefficient on the variable of interest as causal, but not the coefficients on the control variables. After all, doesn't the error term have (by assumption) conditional mean zero? For example, we know that primary school test scores are determined by many factors, some of which are difficult to measure, such as time spent by parents helping with homework and stability of family life. Some of these unmeasured factors are arguably correlated with class size in the United States because of local school funding. To control for these omitted factors, we include socioeconomic indicators, such as district income, in a regression of test scores on class size. Does it make any sense to think of the coefficient on class size as causal, but not the coefficient on district income?

The research on experiments and quasi-experiments has, as a side benefit, spurred more precise thinking about control variables among econometricians. Consider the linear regression model with n observations on a dependent variable Y , a regressor of interest X , a control variable W , and error term u :

$$Y = \beta_0 + \beta_1 X + \beta_2 W + u.$$

The standard textbook assumption justifying ordinary least squares, then and now, is that, given the regressors, the errors have mean zero; that is, $E(u|X, W) = 0$. Under this assumption, the ordinary least squares estimator is unbiased for both coefficients, so both coefficients should have causal interpretations. But as the class

size example makes clear, the control variable W (for example, district income) proxies for deeper unmodeled effects remaining in the error term (for example, parental time spent helping with homework), so the control variable W is correlated with the error term u —indeed, that is why the control variable W is included in the first place. In other words, the key identification assumption in the classical linear regression model, that the error term has a conditional mean of zero, $E(u|X, W) = 0$, does not plausibly apply when W is a control variable.

The experiment/quasi-experiment literature has adopted assumptions that make the distinction between variables of interest and control variables precise and interpretable, and which provide a satisfying answer to our student's hard question. One precise definition of a control variable W is that, once it is included in the regression, the conditional mean of u does not depend on X ; that is, $E(u|X, W) = E(u|W)$. Under this "conditional mean independence" assumption, the coefficient on X is unbiased and consistent, but the coefficient on W is not. Said differently, the coefficient on X can be given a causal interpretation, but the coefficient on W cannot.

Like heteroskedasticity-robust standard errors, clustered standard errors, and some methods for inference with weak instruments, conditional mean independence has made it into modern introductory undergraduate textbooks (for example, Stock and Watson, 2007, pp. 478–80). While the concept of conditional mean independence is not strictly a tool (for example, it is not a command or option in STATA), the shift toward the conditional mean independence assumption has helpfully served to focus attention on measuring a single effect well instead of the vaguer goal of "developing a model of Y ." This assumption also provides a fruitful framework for thinking about regression specification: what control variables do you need so that once you condition on W , it is "as if" X is randomly assigned?

Nonparametric and Semiparametric Methods

Econometricians in the early 1980s knew that parametric (typically linear) functional forms were not always a good approximation. Over the past quarter century, theoretical statisticians and econometricians have worked to develop less restrictive approaches to functional form issues. The result is a well-developed literature on nonparametric and semiparametric estimation. Semiparametric inference is particularly relevant since it focuses on obtaining credible estimates of a specific effect of interest while making very weak subsidiary functional form assumptions (such as how the control variables enter the regression).

Unlike the previous three examples, nonparametric methods have not made it into undergraduate econometrics textbooks. Nonparametric regression is not hard to explain at an intuitive level: kernel regression entails computing a weighted average of Y for observations i with X_i close to some specific point x , thereby estimating $E(Y|X = x)$ without imposing a functional form. But nonparametric methods require large data sets so that there are enough observations close to x to provide a meaningful local average, for all x in the range of the bulk of the data. Although nonparametric and semiparametric methods are

being found increasingly in the program evaluation literature, their use remains specialized and relatively rare.

Identification in Macroeconometrics

I now turn to Angrist and Pischke's challenge that macroeconometricians adopt design-based research. I agree with their suggestion, despite suspecting that it will not take us very far.

Broadly, there are three classes of questions of interest in macroeconometrics. First, why do we observe the specific macroeconomic dynamics that actually occur? This class of questions typically has to do with the estimation of parameters of structural models. Second, what are the effects on macroeconomic dynamics of changes in rules, institutions, and preferences? An example would be to estimate the effect of changing some of those parameters, say from an accommodative set of Taylor rule coefficients for monetary policy to an anti-inflationary set of coefficients, as many have argued happened when Paul Volcker was chairman of the Federal Reserve in the early 1980s.⁴ Finally, what are the effects of autonomous shocks or one-off policy interventions within the context of existing institutions and policy rules? An example of this final class is estimating the effect of an unanticipated policy deviation from the Taylor rule, say an unexpected autonomous increase in the federal funds interest rate by 25 basis points.

Effects of Shocks

Design-based research in the spirit of Angrist and Pischke's essay is arguably best-suited for questions in the third category—that is, questions that investigate the effects of shocks or policy interventions. As an example, consider the current debate over the effect on output of fiscal stimulus. Johnson, Parker, and Souleles (2006) use a quasi-experiment to provide an econometrically clean set of estimates of the dynamic spending pattern, at the individual level, associated with the 2001 one-time federal income tax rebate. Their study exploits random variation in the timing of mailing of rebate checks to disentangle rebate-induced spending changes from macroeconomic factors that affect spending. Specifically, not all the rebate checks could be physically printed and mailed at once, so they were mailed using an algorithm based on the second-to-last digit of Social Security numbers. They find that approximately two-thirds of a rebate check is spent after six months, which is a large fraction from the perspective of the permanent income hypothesis, but still less than the rebate itself. But their study only captures the first round of the multiplier effect, so to speak, which illustrates one limitation of design-based research in macroeconomics—the inability to estimate general equilibrium effects

⁴ A Taylor (1993) rule for monetary policy specifies how a central bank should or does change the nominal interest rate in response to divergences from the bank's target rate of inflation and from potential GDP.

using individual-level data. To estimate general equilibrium effects, one needs known random or as-if random macroeconomic shocks along with data tracking the effect of those shocks on the economy.

A different approach therefore is to use as-if random macro-scale variation in a policy variable—a macroeconomic quasi-experiment—as an instrumental variable. In an application to monetary policy, for example, one can imagine an ideal quasi-experiment: the Federal Open Market Committee decides on a 25 basis point increase, a telex is sent to the New York desk, but there is a typo so the desk instead increases rates by 250 basis points, an error that isn't caught for a month. Suppose these typos occur every now and then, yet are always surprising, so we get repeats of the experiment. Voila, we have a great instrument, the 225 basis point mistakes (unrelated to economic activity, strongly related to the actual federal funds rate)! The “patient,” the U.S. economy, is assumed to be stationary, so data on these repeated quasi-experiments can be used to estimate the same coefficients. Sadly (at least for econometricians), this quasi-experiment is unavailable. In their path-breaking work, Romer and Romer (1989) looked for the next-best thing by reading the minutes of the Federal Open Market Committee to find exogenous variation in monetary policy shocks. This work pioneered the quasi-experiment approach in macroeconometrics by bringing information outside the model to bear on shock identification, but it also remains controversial because of the many detailed judgments that must be made when deciding which monetary policy moves really were exogenous.

Several papers use a quasi-experiment approach to identify dynamic causal effects of fiscal policy shocks. Romer and Romer (2007) use a series derived by reading texts related to tax law changes to construct a measure of exogenous variation. Ramey and Shapiro (1998) and Ramey (2009) focus on government deficit shocks induced by wars: if (as they argue) the wars occur for reasons unrelated to other shocks or economic events, then they are valid instruments for estimation of dynamic causal effects of government spending on output, inflation, and other macroeconomic indicators. The advantage of this approach is that the estimated multipliers incorporate general equilibrium effects. However, this macro quasi-experiment approach is not as clean as its microeconomic counterpart; for example, Ramey (2009) shows that results are sensitive to whether one considers the exogenous variation to be a change in expected future spending or the subsequent spending change itself.

Rigobon's (2003) and Rigobon–Sack's (2003, 2004) scheme for identification of shocks by breaks in variances in a vector autoregression can be viewed as another example of quasi-experiment methods in econometrics: if the variance of the structural shocks change, but the equations linking those shocks to observed macro variables do not, then under certain conditions the coefficients are identified. This idea laudably exploits variation outside of the model for identification, but if the variance changes are small, it shares (in a complicated way) the weak-instrument pitfall of microeconomic design-based research. Moreover, this scheme requires an assumption that in some applications seems heroic (for example, only shock variances changed, not Taylor rule coefficients, in the transition to the Great Moderation).

Another structural vector autoregression identification scheme, a logical extension of Sims's call for using minimal a priori theoretical restrictions, is Faust's (1998) and Uhlig's (2005) idea of obtaining identification by imposing sign restrictions on dynamic causal effects (in the jargon of vector autoregressions, on "impulse response functions"). These restrictions are stunningly minimal. For example, an entire analysis might rely on the assertion that contractionary monetary policy is contractionary, at least in the short to medium run—that is, macroeconomics, despite its shortcomings, gets the sign right, at least at the two-year horizon. One might think that such a minimal restriction would yield little of use, but if it is repeated often enough (applied to many series) this can restrict the family of allowable impulse response functions to a workably small set (for example, see Ahmadi and Uhlig, 2009). This line of research confronts considerable technical hurdles (for example, see Moon, Schorfheide, Granziera, and Lee, 2009), and there are as yet few applications, but it is a creative approach that pushes forward identification in macroeconometrics using a minimum of subsidiary assumptions.

The Effect of Changing Rules and the Estimation of Structural Parameters

Many important macroeconomic policy debates are about rules and institutions, not shocks, and learning the effect on economic performance of changing a rule or institution is at least as important as tracing the dynamic effect of an unexpected shock. For example, should the Fed adopt an explicit inflation-targeting policy? What are the macro consequences of changes in the tax system, or of proposed financial system reforms? Unfortunately, the scope for quasi-experiments in estimating the effect of changing rules or institutions is much more limited than for estimating the effect of shocks. It is hard to imagine a real-world quasi-experiment in which a central bank changed its monetary policy rule for reasons not rooted in prior macroeconomic conditions and expectations of macroeconomic benefits that would flow from the change. This is not to say that we cannot learn from history; in fact, historical studies can usefully incorporate regression analysis to provide concise summaries of relations among multiple macro variables. Bernanke, Laubach, Mishkin, and Posen (1999) provide a good example of an informative historical study of cross-country variation in monetary policy institutions (with a focus on the policy of inflation targeting), along with what happened after countries changed their institutions—in effect, a differences-in-differences design. But we should not be overly optimistic that the coefficients arising from such studies will have a clean causal interpretation.

The role for quasi-experiments in the estimation of structural parameters is probably even more limited. These parameters largely need to be estimated by exploiting time series variation in macroeconomic quantities using model-based identification. Even so, some of the technical developments arising from the reaction to the Sims/Hendry/Leamer critiques can inform and contribute to this work. In particular, estimation of dynamic stochastic general equilibrium models is known to be plagued by complicated versions of the weak instruments problem, and there is active and much-needed work on conditions for identification and on

inference in these kinds of models that is robust to weak instruments (Komunjer and Ng, 2009; Guerron-Quintana, Inoue, and Kilian, 2009).

Concluding Remarks

The developments spurred by the Sims/Hendry/Leamer trio of papers have been important and valuable and have made applied econometric research more credible and influential today than it was 30 years ago. This progress has stemmed both from attention to core identification issues—exploiting credible sources of exogeneity, as in experiments and quasi-experiments—and to technical work in the econometric trenches by theorists laboring to minimize the impact of “whimsical” subsidiary modeling assumptions on statistical inference.

I agree with the theme of Angrist and Pischke that the rise of true experiments in econometrics and a deepened understanding of experimental design and pitfalls has been an unexpected and welcome development. Hendry and Leamer spoke highly of experiments in the early 1980s, but their focus on ways to improve the credibility of observational studies suggests that they did not expect experiments to play much of a role in econometric research. As it turned out, however, public agencies and private companies have proven increasingly willing to support randomized experiments. When well done, experiments elucidate the specific problems under study and also can provide economic lessons beyond the problem at hand.

Looking for sources of credible identification outside the confines of a narrow economic model has also been highly fruitful. An example of a successful quasi-experimental study is Madrian and Shea (2001), who found that 401(k) retirement accounts in which contributions were made unless the participant actively chose to opt-out had much higher take-up rates than 401(k)s in which no contributions were made unless the participants actively chose to opt-in. Indeed, this paper was part of the intellectual case that led to changing U.S. pension laws in the Pension Protection Act of 2006 to encourage default opt-out plans while allowing workers to continue to be able to choose their contributions to such accounts. Their study also provided support for behavioral economic models of procrastination or avoidance of complexity as many workers simply chose the default option, whether it was opt-in or opt-out. Again, when done well, quasi-experiments inform both the specific question under study and the broader corpus of economic knowledge.

I would further suggest that the rise of experiments and quasi-experiments has had a salutary effect on how we teach and evaluate empirical research in economics. The ideas behind internal and external validity have been present in economics for years, but adopting that terminology (which economists did not invent) and tapping into the thinking of statisticians on this topic has helped to organize the conduct and evaluation of empirical work in economics. The presence of some well-done randomized experiments allows economists to compare observational and experimental methods, and I would argue that at least in some cases the observational methods come off looking good. For example, in the question

of how class size affects academic performance, observational regression studies using a rich set of control variables reach quantitatively similar conclusions to those found in the Tennessee Project STAR class size experiment. This suggests that Lalonde's (1986) negative conclusions about observational studies, which he drew after finding that econometric models estimated using observational data on job training programs failed to accord with reliable experimental evidence, should not be over-generalized.

The debate on whether the experimental and quasi-experimental approach in microeconometrics has gone too far is an interesting and important one, and I would have liked to have seen Angrist and Pischke engage it further. There is not space here to recapitulate this debate, but I do encourage interested readers to pursue it by reading the critics, in particular Deaton (2009) and Heckman and Urzua (2009), and the rejoinder to Deaton by Imbens (2009).

■ *I thank Guido Imbens, Michael Kremer, and Mark Watson for helpful conversations. This work was funded in part by NSF grant SBR-0617811.*

References

- Ahmadi, Pooyan Amir, and Harald Uhlig.** 2009. "Measuring the Dynamic Effects of Monetary Policy Shocks: A Bayesian FAVAR Approach with Sign Restrictions." <http://amir-ahmadi.de/Amir%20Ahmadi%20and%20Uhlig%20%282009%29.pdf>.
- Arrelano, Manuel.** 1987. "Computing Robust Standard Errors for Within-Groups Estimators." *Oxford Bulletin of Economics and Statistics*, 49(4): 431–34.
- Arrelano, Manuel.** 2003. *Panel Data Econometrics*. Advanced Texts in Econometrics. Oxford, UK: Oxford University Press.
- Bernanke, Ben S., Thomas Laubach, Frederic S. Mishkin, and Adam Posen.** 1999. *Inflation Targeting: Lessons from the International Experience*. Princeton University Press.
- Bertrand, Marianne, Esther Dufo, and Sendhil Mullainathan.** 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics*, 119(1): 249–75.
- Deaton, Angus.** 2009. "Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development." NBER Working Paper 14690.
- Faust, Jon.** 1998. "The Robustness of Identified VAR Conclusions about Money." *Carnegie-Rochester Series on Public Policy*, vol. 49, pp. 207–244.
- Granger, Clive W. J.** 2003. "Time Series Analysis, Cointegration, and Applications." Nobel Lecture, December 8. http://nobelprize.org/nobel_prizes/economics/laureates/2003/granger-lecture.pdf.
- Guerron-Quintana, Pablo, Atsushi Inoue, and Lutz Kilian.** 2009. "Frequentist Inference in Weakly Identified DSGE Models." Unpublished paper, University of Michigan, October. <http://www-personal.umich.edu/~lkilian/gik102509.pdf>.
- Gujarati, Damodar.** 1978. *Basic Econometrics*, 1st ed. New York: McGraw-Hill.
- Hansen, Christian.** 2007. "Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data when T is Large." *Journal of Econometrics*, 141(2): 597–620.
- Heckman, James J., and Sergio Urzua.** 2009. "Comparing IV with Structural Models: What Simple IV Can and Cannot Identify." NBER Working Paper 14706.
- Hendry, David F.** 1980. "Econometrics—Alchemy or Science?" *Economica*, 47(188): 387–406.

- Ibragimov, Rustam, and Ulrich Müller.** Forthcoming. "t-statistic Based Correlation and Heterogeneity Robust Inference." *Journal of Business & Economic Statistics*.
- Imbens, Guido W.** 2009. "Better LATE than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)." NBER Working Paper 14896.
- Johnson, David S., Jonathan A. Parker, and Nicholas S. Souleles.** 2006. "Household Expenditure and the Income Tax Rebates of 2001." *American Economic Review*, 96(5): 1589–1610.
- Kiefer, Nicholas M.** 1980. "Estimation of Fixed Effects Models for Time Series of Cross-Sections with Arbitrary Intertemporal Covariance." *Journal of Econometrics*, 14(2): 195–202.
- Komunjer, Ivana, and Serena Ng.** 2009. "Dynamic Identification of DSGE Models." <http://econ.ucsd.edu/~ikomunje/files/papers/spectral.pdf>.
- Lalonde, Robert John.** 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *American Economic Review*, 76(4): 604–20.
- Leamer, Edward E.** 1983. "Let's Take the Con Out of Econometrics." *American Economic Review*, 73(1): 31–43.
- Madrian, Brigitte C., and Dennis F. Shea.** 2001. "The Power of Suggestion: Inertia in 4501(k) Participation and Savings Behavior." *Quarterly Journal of Economics*, 116(4): 1149–87.
- Moon, Hyungsik Roger, Frank Schorfheide, Eleonora Granziera, and Mihye Lee.** 2009. "Inference for VARs Identified with Sign Restrictions." <http://www.econ.upenn.edu/~schorf/papers/svar-paper.pdf>.
- Moreira, Marcelo J.** 2003. "A Conditional Likelihood Ratio Test for Structural Models." *Econometrica*, 71(4): 1027–48.
- Pindyck, Robert S., and Daniel L. Rubinfeld.** 1981. *Econometric Models and Economic Forecasts*, 2nd ed. New York: McGraw-Hill.
- Ramey, Valerie A.** 2009. "Identifying Government Spending Shocks: It's All in the Timing." NBER Working Paper 15464.
- Ramey, Valerie A., and Michael D. Shapiro.** 1998. "Costly Capital Reallocation and the Effects of Government Spending." *Carnegie Rochester Conference on Public Policy*, vol. 48, pp 145–209.
- Rigobon, Roberto.** 2003. "Identification through Heteroskedasticity." *Review of Economics and Statistics*, 85(4): 777–92.
- Rigobon, Roberto, and Brian Sack.** 2003. "Measuring the Reaction of Monetary Policy to the Stock Market." *Quarterly Journal of Economics*, 118(2): 639–69.
- Rigobon, Roberto, and Brian Sack.** 2004. "The Impact of Monetary Policy on Asset Prices." *Journal of Monetary Economics*, 51(8): 1553–75.
- Romer, Christina D., and David H. Romer.** 1989. "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz." *NBER Macroeconomics Annual*, vol. 4, pp. 121–170.
- Romer, Christina D., and David H. Romer.** 2007. "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks." NBER Working Paper 13264.
- Sims, Christopher A.** 1980. "Macroeconomics and Reality." *Econometrica*, 48(1):1–48.
- Stock, James H., and Mark W. Watson.** 2007. *Introduction to Econometrics*, 2nd ed. Boston: Addison-Wesley.
- Taylor, John B.** 1993. "Discretion versus Policy Rules in Practice." *Carnegie-Rochester Conference Series on Public Policy*, vol. 39, pp.195–21.
- Uhlig, Harald.** 2005. "What Are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure." *Journal of Monetary Economics*, 52(2): 381–419.
- White, Halbert.** 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica*, 48(4): 817–38.
- Wonnacott, Ronald J., and Thomas H. Wonnacott.** 1979. *Econometrics*, 2nd ed. New York: Wiley.