Identification Is Not Causality, and Vice Versa

R. Kahn
University of Michigan

Toni M. Whited
University of Michigan and NBER

We distinguish between identification and establishing causality. Identification means forming a unique mapping from features of data to quantities that are of interest to economists. Establishing causality by finding sources of exogenous variation is often considered synonymous with identification, but these two concepts are distinct. Exogenous variation is only sometimes necessary and never sufficient to identify economically interesting parameters. Instead, even for causal questions, identification must rest on an underlying economic model. We illustrate these points by analyzing identification in three recent papers and by examining the estimation of a simple dynamic model. (*JEL* C21, C26, G31, G32)

Received June 6, 2017; editorial decision September 26, 2017 by Editor Gregor Matvos. Authors have furnished supplementary code, which is available on the Oxford University Press Web site next to the link to the final published paper online.

In terms of its pure statistical definition, identification is simple. An applied econometrician defines an objective function over parameters and a data population, and her goal is to select parameters that minimize this objective function, in which the population has been replaced by a specific sample. A parameter is identified if there is a unique minimum for the objective function at its true value in the population.[1] Yet discussion

This work would have been impossible without the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575. We thank Shuangning Zhu for excellent research assistance and Gregor Matvos, Luke Taylor, and Missaka Warusawitharana for helpful comments. The material in this paper draws in part from Kahn and Whited (2016), with permission from Now Publishers. Send correspondence to Toni M. Whited, Stephen M. Ross School of Business, University of Michigan, Ann Arbor, MI 48109-1234; telephone: 734-764-1269. E-mail: twhited@umich.edu.

[1] We limit our discussion to point identification instead of partial or set identification (Chernozhukov, Hong, and Tamer 2007).

© The Author 2017. Published by Oxford University Press on behalf of The Society for Financial Studies. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com.
doi:10.1093/rcfs/cfx020 Advance Access publication November 3, 2017

Downloaded from https://academic.oup.com/rcfs/article-abstract/7/1/1/4590088
by Wuhan University Library user
on 02 March 2018

of this pure statistical issue of identification is not of particular interest to applied economists because the parameter at the minimum of the objective function may or may not be of interest from an economic point of view. For example, if we say a regression of price on quantity does not identify demand, we are not arguing that the regression itself is not well formed. Ordinary least squares (OLS) produces an unbiased estimate of the slope coefficient on price. However, we are stating that this estimation has not identified an economic parameter, typically a utility parameter, that we find interesting. The true problem of identification is then using an econometric objective function to form a mapping from observed data to relevant economic parameters. Unfortunately, identifying an economically interesting parameter is far more difficult than the sheer statistical definition of identification might suggest.

The purpose of this paper is to delineate the relationship between estimating a causal effect and the more general issue of identification. At least since Alfred Marshall's *Principles of Economics*, economists have understood causality as a ceteris paribus comparison: the causal effect of variable A on variable B is the change in B that results from altering A while holding all other features of the world constant.[2] Causal effects are simply elasticities, but they are difficult to estimate because econometricians rarely observe occasions where one variable is altered while others are held constant, that is, where there is genuine exogenous variation in a variable.

This exogenous variation forms the focus of how many economists think about identification. However, the general issue of identification is broader in scope than the establishment of exogenous variation. We elaborate on this point with the main observation that identification relates to parameters of a model. In many applications, causation is the ultimate question of interest, so these parameters are the elasticities that define causal effects. In these instances, the model is an econometric model, typically a regression. In such a situation, it is common to see researchers adopt a quasi-experimental approach. For example, economists are often interested in the effect of limited government interventions on economic variables: how an increase in the minimum wage affects employment (Neumark and Wascher 1992; Card and Krueger 1994; Dube, Lester, and Reich 2010; Sorkin 2015), how class size affects achievement (Angrist and Lavy 1999; Krueger 1999; Krueger and Whitmore 2001; Chetty et al. 2011), or how training affects earnings (Ashenfelter 1978; Ashenfelter and Card 1985; Heckman, Ichimura, and Todd 1997). In all of these cases, the parameter of interest is a simple elasticity. The government has a lever at its disposal, and the question is the outcome when the government pulls that lever. This simple ceteris

[2] See, for instance, the discussion in Heckman and Pinto (2015).

Downloaded from https://academic.oup.com/rcfs/article-abstract/7/1/1/4590088
by Wuhan University Library user
on 02 March 2018

paribus comparison surrounding a specific and limited government intervention makes identification relatively straightforward. If the goal of a study is simply to establish the average effect of a previous intervention, then as long as this intervention is plausibly exogenous, the causal link to a government policy has been identified. Yet even in this straightforward context, the average treatment effect that comes from such an approach is limited in its applicability. It represents only an estimate of the average causal effect of a variable under a particular, historical intervention. While understanding historical interventions is often of interest, without additional assumptions, it is difficult to extrapolate any such results to predictions about future interventions of a similar type.

The tight connection between causality and identification in these popular quasi-experimental studies makes it easy to confuse identification with the establishment of causality through exogenous variation. In fact, Angrist and Pischke (2008) present the issue of identification entirely as a search for an approximation to an ideal experiment. However, not all questions of interest can be phrased in experimental terms. In particular, in corporate finance, we are rarely confronted with the strong policy levers that have made the estimation of treatment effects one of the central activities of many areas of applied microeconomics.

However, not all interesting questions are causal in nature, and not all identification issues revolve around establishing causality, so the second point we wish to make is that sometimes the model that provides identification is an economic model. For example, one might be able to run an experiment to establish that a causal effect exists, but the experiment alone typically cannot identify the economic forces that are behind the causal effect, and these forces are usually at least as interesting as the effect itself. For example, Breza (2012) estimates the causal effect of peer repayment on individuals' repayment decisions using plausibly exogenous variation in loan maturity surrounding a default crisis in India. However, to understand whether positive or negative influence is more important for the operation of the peer effects, the paper structurally estimates the parameters of a dynamic discrete choice model, finding that positive influences are more important.

Several simple corollaries follow from our second point. Finding exogenous variation in a variable is never sufficient for identification of an economically interesting parameter, as identification is always based on a verbal or mathematical theory. Thus, identification can never be free of assumptions or even light on assumptions. The necessity of assumptions means that for some questions and some types of structural estimation, exogenous variation may not even be necessary for identification.

Our final point is that neither the presence of random variation nor the establishment of causality necessarily fulfills the goal of answering

Downloaded from https://academic.oup.com/rcfs/article-abstract/7/1/1/4590088
by Wuhan University Library user
on 02 March 2018

interesting questions. None of these points are entirely new. In fact, the last can be traced back at least as far as Koopmans (1949), who pointed out:

> Where statistical data are used as one of the foundation stones on which the equation system is erected, the modern methods of statistical inference are an indispensable instrument. However, without economic "theory" as another foundation stone, it is impossible to make such statistical inference apply directly to the equations of economic behavior which are most relevant to analysis and to policy discussion.

We illustrate these points in two ways. First, we see how three different papers identify an economic parameter. Along the way, we also relate each of these identification strategies back to the statistical definition of identification.

Second, we explore the identification and estimation of a simple, canonical dynamic leverage model. This exercise allows us to make two further points. First, we draw an important analogy between reduced-form and structural estimation. We demonstrate that the search for mechanisms that underlie the causal elasticities in reduced-form studies is strongly related to the identification of model parameters in a structural estimation. Both undertakings involve establishing mappings from model predictions to model parameters, with the difference between the two types of estimation being the use of verbal versus mathematical models. Second, we use a Monte Carlo simulation of an indirect inference estimator to show that exogenous variation may be superfluous even when it is available. The reason is multiple mappings from a single causal elasticity to various model parameters, with the result that the causal elasticity provides insufficient identifying information for any single parameter. We close by arguing that those instances in which exogenous variation is useful for identification of economic parameters are those in which a unique mapping exists between a causal elasticity and an interesting economic parameter. In those cases, one would naturally prefer a reduced-form approach to estimation in the first place.

1. Three Papers

1.1 Identification with exogenous variation

Our first example is Bennedsen et al. (2007), which studies Danish family firms, asking whether in-family succession of CEOs hurts performance. Given that most firms in the world are family firms, this question is clearly of interest. This question can also be phrased as a ceteris paribus comparison: how would the performance of the company have been

Downloaded from https://academic.oup.com/rcfs/article-abstract/7/1/1/4590088
by Wuhan University Library user
on 02 March 2018

different if an outside hire had been chosen as a CEO instead of a family member? In contrast to ceteris paribus comparisons involving the impact of a government policy, here the parameter of interest is not an observable elasticity but a summary measure of an underlying agency problem. If an in-family CEO is appointed, he is drawn from a limited pool of family members. Within this limited pool, candidates are unlikely to be as proficient as if they were drawn from the broader market outside the family. So while the inside hire creates a nonpecuniary benefit to the family, it could hurt the performance of the firm. To understand the magnitude of this agency problem, we want to estimate the loss in performance that is due to the choice of a family member over an outsider. The loss in performance is a consequence caused by the choice of a family member over an outside candidate, but the parameter that measures the performance loss is of interest not because it represents an average, presumably causal, estimated effect: it is interesting because it represents a deeper agency friction.

Identifying the parameter that represents this agency friction is difficult because demand for a family CEO is endogenous to the performance of the company. Poor performance may force the family to choose an outside CEO instead of a relative, and good performance may make a family insouciant about the specter of an incompetent family CEO. In this case, in a simple regression of performance on the choice of CEO:

$$\text{performance} = \alpha + \beta(\text{In–family succession}) + u,$$

the coefficient on in-family succession, β, does not identify agency costs, because it is a function of both agency costs and the unobserved economic variables affecting demand for family CEOs. While the statistical parameter is well defined, the OLS objective function does not have a unique minimum at the true value of agency costs, as the model of the underlying economics of the question does not allow for a mapping of this regression slope, β, to the agency parameter of interest.

To identify the agency friction, the econometric model needs more structure. With data from Danish firms, the authors can observe the demographics of controlling families. They choose the gender of the firstborn child as an instrument to determine the causal impact of in-family succession. The argument the authors present is that families without male firstborn children will be less likely to choose a family CEO, yet families with and without male firstborn children should have ex-ante identical performance, as the biology of child gender is genuinely random.

Crucially, while biology buys randomness, the power of the instrument in identifying the agency frictions comes from two additional assumptions added to the model. For the agency parameter to be identified, the reader has to believe the following. First, female CEOs will be no

Downloaded from https://academic.oup.com/rcfs/article-abstract/7/1/1/4590088
by Wuhan University Library user
on 02 March 2018

different from male CEOs, otherwise the instrument does not satisfy the exclusion restriction. Second, Danish families have a preference for primogeniture, otherwise the instrument will be weak (Staiger and Stock 1997). If both of these assumptions hold, then the instrumental variables objective function formed from the gender of the eldest child has a unique minimum at the agency cost friction, which is then identified.

These identifying assumptions are relatively mild, but they are still assumptions, and family preference for primogeniture is less innocuous than it seems. To see the importance of this assumption, note that most completely random variables, for example, the inches of rainfall in Kansas in a year, are useless as instruments for identifying the effects of agency problems on firm performance because Danish family firms do not react to these irrelevant sources of randomness. The gender of a firstborn child makes a good instrument not only because it is random but because Danish firms react to the instrument. Thus, in order to use the exogenous variation from the gender of the firstborn, we must also assume controlling families are somewhat sexist. In the absence of sexism, the instrument has no bearing on the succession decision, and the parameter is again unidentified. If we are willing to assume that sexism exists, then the exclusion restriction is that sexism affects firm performance only through the choice of a family CEO, and this assumption is nontrivial. For instance, imagine that sexism causes controlling families to raise firstborn boys and girls differently. Boys are groomed to lead the family firm, while girls are encouraged to pursue other professions. In this case, the gender of the (potential) family CEO would affect (potential) firm performance, and the exclusion restriction would not hold. This issue muddies the interpretation of the results, as firstborn girls who do become CEOs are likely to be particularly talented if the grooming story is true. Thus, the negative relation between in-family CEO succession and performance would in part be due to the high-performing girls instead of the low-performing boys, and the verbal model that identifies agency issues centers around the low-performing boys. The data provide no evidence for or against the possibility of grooming, so identification requires that one assume away this possibility. This example thus illustrates that exogenous variation alone does not allow identification of an interesting economic parameter, the agency parameter. Instead, identification comes from the combination of exogenous variation and the assumptions made in the paper's verbal theory of behavior.

1.2 Identification without exogenous variation

For some questions, exogenous variation is not even necessary to identify an economically interesting parameter. In fact, causal inference in general may not be the point. To illustrate this point, we turn outside

Downloaded from https://academic.oup.com/rcfs/article-abstract/7/1/1/4590088
by Wuhan University Library user
on 02 March 2018

corporate finance and examine Davis, Fisher, and Whited (2014), which asks the extent to which agglomeration externalities affect aggregate consumption growth. Agglomeration externalities are the productivity gains that occur when workers and firms locate in the same area. In this case, the identification of the impact of agglomeration on consumption growth is more difficult, in large part because the question cannot be phrased as a ceteris paribus comparison. For example, one would want to compare consumption growth in Chicago as it currently is with consumption growth in a counterfactual "city" in which Chicago's population is spread out over Illinois (but no other changes are made). Of course, such a situation is difficult to envision, and impossible to observe.

Nonetheless, the process used to establish identification carries many similarities with the process used to establish identification in Bennedsen et al. (2007). Davis, Fisher, and Whited (2014) starts with an explicit set of assumptions that underlie a dynamic general-equilibrium model of agglomeration in cities. In the model, firms do not take into account the positive productivity spillovers that they generate when they hire extra workers within a city. This externality then affects consumption growth as long as land prices are rising. In this case, an increase in the forecast of land prices leads firms to economize on space now. This reaction in turn leads to an increase in productivity via the externality because more workers and firms are clustered onto a smaller space. The set of assumptions that leads to this behavior in the model in turn implies that the correlation between a forecast of land prices and the growth of total factor productivity is a function of the agglomeration externality and some other easily estimated parameters. This result means in turn that the OLS objective function in a regression of total factor productivity on forecasted land prices has a unique minimum at the true value of the agglomeration externality parameter, without the need for any exogenous source of variation.

This identification is not assumed. Just as in Bennedsen et al. (2007), it is the result of a careful argument extended from a set of assumptions. However, there are two important differences between identification in the structural study and identification in the reduced-form study. First, the arguments in Davis, Fisher, and Whited (2014) are phrased using mathematics, and the arguments in Bennedsen et al. (2007) are verbal. Second, all of the assumptions needed for identification are contained in Davis, Fisher, and Whited (2014). In contrast, some of the identifying assumptions in Bennedsen et al. (2007) are not contained in the paper, even though the paper is quite explicit in stating that the gender of the firstborn can only affect CEO succession via the choice of a family CEO, and even though the paper is as careful as it can be to convince the reader that all possibilities for violation of the exclusion restriction have been exhausted. Despite the cleanness of the natural experiment and the high level of care taken in its execution, because the identifying assumptions

Downloaded from https://academic.oup.com/rcfs/article-abstract/7/1/1/4590088
by Wuhan University Library user
on 02 March 2018

are verbal, there is always room to consider new alternatives, such as the priming we describe above, that could violate the exclusion restriction.

1.3 Using models and exogenous variation

By now it is clear that using both exogenous variation and models requires assumptions. As such, each method is imperfect, so the next natural question to ask is whether these two methods can be used in conjunction to hone inference. To illustrate this point, we look at model estimation and the natural experiment in Li, Whited, and Wu (2016). The question is whether collateral matters for leverage.

The model is a partial-equilibrium, dynamic agency model that produces a collateral constraint as the endogenous outcome of a contracting problem. The parameters of the model are estimated by matching model-generated moments with moments computed from actual data. Technically, identification requires that the mapping from the model parameters to the moments be one-to-one and onto. From an intuitive perspective, this identification condition means that the relations between the moments and parameters need to be steep and monotonic, that is, the moments need to be informative about the parameters. Steep, monotonic relations result in an econometric objective function that has a unique minimum, so the parameters are identified. For example, in the model in Li, Whited, and Wu (2016), the unobservable variance of a productivity shock is closely related to the variance of observable average operating profits, so the unobservable variance parameter is identified from an econometric objective function constructed in part by setting the variance of operating profits in a data set equal to the model-implied variance of operating profits.

The identifying assumptions themselves are precisely the assumptions that define the structure of the model. While these assumptions are not testable, parameter identification typically manifests itself in tight parameter standard errors (Bazdresch, Kahn, and Whited forthcoming). Of particular interest for our analysis of identification is the result that the main model parameter, the one that quantifies the collateral value of assets, is precisely estimated and thus likely well identified.

However, the question remains whether too many relevant forces have been left out of the model and whether these omitted forces load onto the collateral parameter in such a way that it captures not only the collateral value of assets but perhaps any such omitted forces. In this case, the model parameter embodies many forces besides the actual quantity of interest, which is the collateral value of assets. Thus this collateral value is not identified by the estimated model parameter.

To tackle this question, the paper challenges the external validity of the model using a natural experiment on the value of collateral. The setting is secured lending through a special purpose vehicle (SPV). When

Downloaded from https://academic.oup.com/rcfs/article-abstract/7/1/1/4590088
by Wuhan University Library user
on 02 March 2018

borrowing via an SPV, the firm sells the collateral to the SPV, which is then exempt from the automatic stay in bankruptcy. Thus, it is easier for the lender to seize the collateral than it would be if the firm were to keep the collateral on its books. However, a bankruptcy judge can have the leeway to recharacterize the sale as a loan, at which point, the collateral goes back to the firm and is subject to the automatic stay. This discretion introduces substantial uncertainty about the value of collateral in the event of seizure.

The natural experiment revolves around two sets of events that affect judge discretion. The first event is a set of anti-recharacterization laws that were passed at the state level, specifically, in Texas and Louisiana in 1997 and in Alabama in 2001. Because the lobbying for these laws came neither from the banks that lend to industrial firms nor from the industrial firms themselves, these laws can be thought of as plausibly exogenous from a political economy perspective. The second event is an important court case in 2003 that nullified the laws. Specifically, in the *Reaves Brokerage Company, Inc. v. Sunbelt Fruit & Vegetable Company, Inc.*, the judge ignored the Texas statute, applied federal law, and recharacterized the collateral. In the seven years following this court case, it served as a cited precedent in sixty-two other bankruptcy cases.

These events lend themselves to a standard difference-in-difference estimation. The first dimension of the diff-in-diff is incorporation in Texas, Louisiana, or Alabama, which defines a set of treated firms. The second dimension is the time in which the laws were in effect, defined as after the law in the relevant state was passed but before 2004. A standard regression analysis finds that the laws have a statistically significant effect of 0.04 on leverage, whose average level in the sample is approximately 0.3. Thus, the effect is sizable.

The next question that comes to mind is whether the collateral parameter in the model changes when the model is estimated on data from the four different samples that define the natural experiment: the treated and control firms, with and without the laws in effect. If the main model parameter really embodies the value of collateral, then the estimated value of this parameter should change in the treated group but not in the control group. The answer is largely yes. The value of this parameter rises by approximately 0.052 in the group of treated firms when the laws are in effect. This change is statistically significant and nearly the same magnitude as the simpler regression-based estimate, so the experiment validates the model. However, the collateral parameter drops by about 0.026 in the control group, so the final difference-in-difference estimate obtained from the model is nearly 0.08. Thus, although the sign and statistical significance of the structural difference-in-difference estimate is the same as the reduced-form difference-in-difference estimation, the magnitude is different.

Downloaded from https://academic.oup.com/rcfs/article-abstract/7/1/1/4590088
by Wuhan University Library user
on 02 March 2018

This last result brings up the second point we wish to make in this example. Models and natural experiments can be used together to achieve identification that is more convincing than either could achieve in isolation, as both require often strong assumptions. In the anti-recharacterization experiment, the difference-in-difference regression strictly gives identification only of the elasticity of leverage to the enactment of the laws. As we have emphasized earlier, if this elasticity is the ultimate object of interest, then this level of identification is sufficient. However, the regression does not identify the causal effect of a change in the value of collateral if unobservable forces changed at the same time.

At this point, the model becomes useful because one can use modeling assumptions to control for at least a subset of possible unobservables. The relevant unobservable in this setting is demand uncertainty, which is notoriously hard to estimate at the firm level and therefore nearly impossible to include as a control in a difference-in-difference regression. However, demand uncertainty can easily be captured by a parameter in a model, so it is possible to recover average uncertainty in the treated and control firms, with and without the laws. While uncertainty is nearly unchanged in the treated group, with and without the laws, uncertainty is markedly higher in the control group without the laws than in the control group with the laws. Thus, even though leverage is identical in these last two groups, the control group, which faces higher uncertainty, optimally conserves debt capacity. This behavior implies that the value of collateral is much higher in the control group without the laws. Thus, the structural difference-in-difference effect on collateral is larger than the reduced-form elasticity.

In the end, the natural experiment is useful as an external validity check on the model. In turn, the model helps isolate the value of collateral, which is embedded in the reduced-form difference-in-difference elasticity, but which is not completely identified by this elasticity because of the presence of concurrent changes in uncertainty.

2. Dynamic Model

In this section, we sketch a simple dynamic model to illustrate two related points. First, elasticities can be a function of many interesting economic parameters, even when these elasticities are driven by truly exogenous variation. We use this observation to argue that exercises in identifying economic parameters in structural estimations are highly analogous to attempts to uncover the mechanisms that underlie reduced-form quasi-experimental elasticities. Second, while we have already argued that exogenous variation is not necessary to identify interesting economic parameters, we show that it does not necessarily aid in the identification of parameters in a structural estimation even when it is available. We

Downloaded from https://academic.oup.com/rcfs/article-abstract/7/1/1/4590088
by Wuhan University Library user
on 02 March 2018

then argue that in those instances when exogenous variation is helpful, structural estimation is likely not even necessary.

2.1 Assumptions

The model is the simple textbook setup from Strebulaev and Whited (2012), which is an infinite-horizon partial equilibrium model of the firm cast in discrete time. Each period, a risk-neutral manager, acting on behalf of shareholders, chooses the capital stock to maximize the value of the firm, which is the expected present value of the stream of future cash flows to (or from) shareholders. These cash flows are simply operating profits minus investment expenditures. We assume a stochastic, decreasing returns-to-scale profit function, zk^{α}, $0 < \alpha < 1$, in which z is an exogenous demand or technology shock, k is the capital stock, and α is a returns-to-scale parameter. Investment is given via the identity:

$$I \equiv k' - (1-\delta)k, \tag{1}$$

in which a prime indicates a variable in the subsequent period, and in which δ is the capital depreciation rate. Given this notation, current-period disributions to shareholders are given by.

$$e(z,k,k') \equiv \; zk^{\alpha} - k' + (1-\delta)k. \tag{2}$$

Next, we assume that the shock z follows a specific Markov process—a first-order autoregressive process, $AR(1)$, in logs:

$$\ln(z') = \rho \ln(z) + \sigma\varepsilon'. \tag{3}$$

Here, ε is an independently and identically distributed (i.i.d.) standard normal random variable, ρ is an autoregressive coefficient, and σ is a variance parameter.

Thus far, the model contains no financing frictions. As is by now standard, we add a simple, reduced-form financing friction that can be interpreted as the cost of issuing new external finance (Gomes 2001; Hennessy and Whited 2005). External finance is necessary in those states of the world in which $e(z,k,k') < 0$. If $e(z,k,k') > 0$, the firm simply distributes these proceeds to shareholders. However, if $e(z,k,k') < 0$, the firm extracts this amount from its shareholders and pays a proportional fee $\lambda e(z,k,k')$. This situation closely resembles a rights issue and can thus be thought of as an equity issuance.

We can now express the dynamic optimization problem via the Bellman equation:

$$V(k,z) = \max_{k'} \left\{ (zk^{\alpha} - k' + (1-\delta)k \;)(1 + \lambda \mathbb{I}_{e<0}) + \frac{1}{1+r} \mathbb{E} V(k',z') \right\}. \tag{4}$$

Downloaded from https://academic.oup.com/rcfs/article-abstract/7/1/1/4590088
by Wuhan University Library user
on 02 March 2018

Here, $\mathbb{I}_{e<0}$ is an indicator function for those states of the world in which the firm is issuing equity, r is the risk-free rate of interest, and $\mathbb{E}$ is the expectation operator with respect to the conditional distribution of z implied by Equation (3).

2.2 Optimality Conditions and Comparative Statics

If we differentiate Equation (4) with respect to k', and use the Envelope condition, we get the familiar result that the expected marginal product of capital net of depreciation equals the interest rate:

$$\mathbb{E}((\alpha z' k'^{\alpha-1} + (1-\delta))(1+\lambda \mathbb{I}_{e<0})) = r. \tag{5}$$

To see how the problem of identification arises in this model, it is useful to simplify the model to obtain a closed form for Equation (5). Specifically, if we set the cost of equity issuance, λ, equal to zero, we can use the lognormality of z to rewrite (5) as:

$$\exp\left(\rho \ln(z) + \frac{1}{2}\sigma^2\right)\alpha k'^{\alpha-1} = r + \delta. \tag{6}$$

Simple inspection of the first-order condition in Equation (6) shows that the response of investment, k'/k, to the exogenous profitability shock, z, depends on all of the model parameters. Thus, even if one could observe z perfectly, estimating the genuinely causal elasticity of investment with respect to z would yield insufficient information to identify any of the model parameters. Although there is no closed-form expression for Equation (5) in terms of the model parameters, it is clear that a similar issue arises in this more complex setting.

To illustrate this problem further, we consider a simple parameterization of the full ($\lambda > 0$) model and conduct a comparative statics exercise in which we examine how the model parameters affect the sensitivity of investment, $I\ /\ k$, to the exogenous component of profits, z. Here, it is important to note that we can interpret this sensitivity as a causal effect in the laboratory of this specific model because z is modeled as exogenous.

This model has five parameters: the profit function curvature, α; the standard deviation and serial correlation of the shock, σ and ρ; the depreciation rate, δ; and the cost of equity issuance λ. For simplicity, we set the parameters to be close to those from recent papers that consider similar models (e.g., Glover and Levine, forthcoming; Bazdresch, Kahn, and Whited, forthcoming). Specifically, we set $\alpha = 0.7$, $\rho = 0.7$, $\sigma = 0.2$, $\delta = 0.1$, and $\lambda = 0.05$.[3]

[3] Another model parameter is the interest rate, which is typically estimated with easily available aggregate data, and which is therefore separately identified in most applied applications. We set $r = 0.04$.

Downloaded from https://academic.oup.com/rcfs/article-abstract/7/1/1/4590088
by Wuhan University Library user
on 02 March 2018

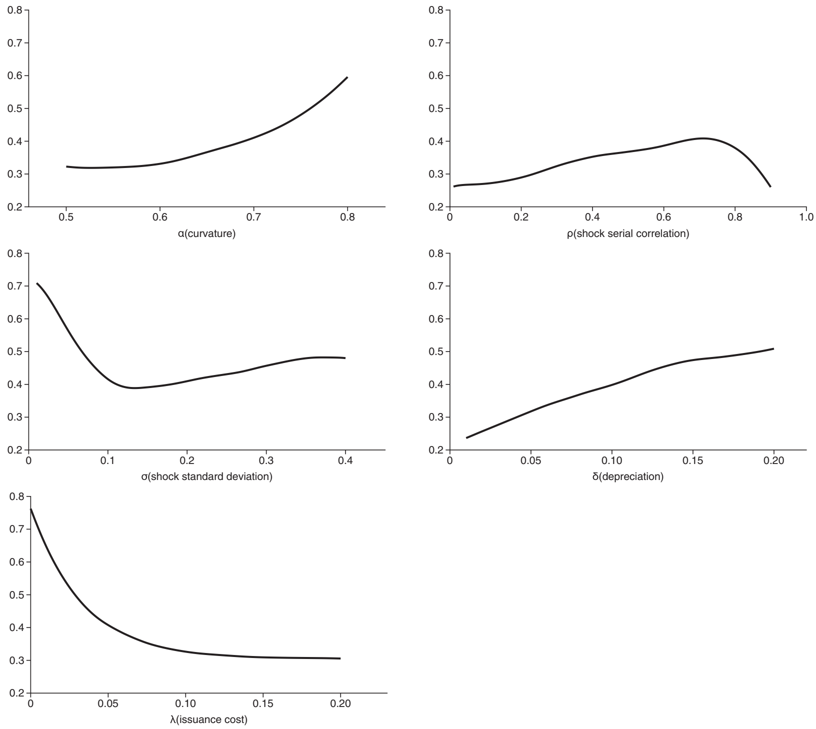

Figure 1
Causal elasticicities and model parameters.
The y-axis on each plot is the slope coefficient from regressing investment, $I\ /\ k$, on the exogenous profitability shock, z. On each x-axis is a different parameter from the model in Section 2.

Figure 1 presents our comparative statics exercises. Each plot shows the sensitivity of investment to z as a function of a specific model parameter. The most salient result in Figure 1 is that all model parameters affect the sensitivity of investment to z. Interestingly, as in Moyen (2004) and Hennessy and Whited (2007), this sensitivity is declining in the cost of external finance, as the firm optimally responds less aggressively to shocks when external finance is expensive. In addition, the relations between this sensitivity and both ρ and σ are nonmonotonic. The reason is that the sensitivity of investment to z is a ratio. The relation between this sensitivity and the covariance term in the numerator is monotonically increasing in ρ and σ, while relation between this sensitivity and the variance term in the denominator is monotonically decreasing in ρ and σ. Thus, the net effect is ambiguous. In the end, the important takeaway from Figure 1 is that the various model parameters and the

Downloaded from https://academic.oup.com/rcfs/article-abstract/7/1/1/4590088
by Wuhan University Library user
on 02 March 2018

corresponding economic mechanisms that influence even a causal relation can be many and hard to identify.

2.3 Mechanisms, channels, and identification

This issue is highly familiar to reduced-form researchers. For example, Bernstein (2015) estimates the causal effect of going public on innovation, using NASDAQ returns between initial public offering (IPO) filing and IPO completion as an instrument. However, estimating this arguably causal elasticity is not informative about the channel whereby IPOs hurt innovation, so the paper performs several auxiliary analyses to attempt to identify the channel. These sorts of channels or mechanisms are analogous to parameters in a model, as the models in most reduced-form studies are verbal. For example, Bernstein (2015) considers, and ultimately rejects, the possibility that private firms suffer from agency problems that might engender excessive innovation before the IPO. The auxiliary test is a comparison of private firms that are professionally managed by venture capitalists with private firms that are managed by entrepreneurs. The verbal model supporting this comparison is a simple agency model in which entrepreneurs extract private benefits from excessive innovation but professional managers do not, so, on a conceptual level, a private benefit parameter is smaller for professional managers. Thus, the purpose of the auxiliary test is to identify the private benefit parameter, which can then help rule out an excessive-innovation hypothesis.

This type of auxiliary analysis has a strong analogy in structural estimation, in which one often identifies model parameters via different model predictions, which typically take the form of the signs and magnitudes of statistical quantities such as means, variances, and regression coefficients. As described in Section 1.3, parameter identification requires a steep and monotonic relation between a model prediction and a model parameter, with the result that an econometric objective function based on such a prediction has a unique minimum in the dimension of the parameter in question. For example, the incidence of equity issuance is monotonically decreasing in the cost of equity issuance, λ, so setting issuance incidence in the data equal to model-implied issuance incidence helps identify λ.[4]

Each of these correspondences between model predictions and parameters is strongly akin to auxiliary tests run in quasi-experimental corporate finance studies, which are designed to identify the economic mechanisms that underlie any causal relations. Both of these exercises

[4] See Andrews, Gentzkow, and Shapiro (2017) for a formal measure of the informativeness of moments.

Downloaded from https://academic.oup.com/rcfs/article-abstract/7/1/1/4590088
by Wuhan University Library user
on 02 March 2018

tie a typically endogenous pattern in the data to the parameter of either a verbal or mathematical model. This type of identification exercise in structural estimation has both attractive and unattractive properties. On the one hand, the correspondences between model predictions and model parameters are typically intuitive, and all of the assumptions behind these correspondences are explicit. On the other hand, these exercises lean heavily on the structure and assumptions that underlie the model. Auxiliary tests also have both attractive and unattractive properties. While often intuitive, these types of tests are typically based on either verbal or implicit assumptions, whose plausibility is thus sometimes difficult to assess.

3. Exogenous Variation in Structural Estimation

This discussion of parameter identification in structural estimation raises the interesting question of whether exogenous variation is useful for identifying model parameters. While we have already shown that one can interface natural experiments with structural estimations, this demonstration does not shed light on whether exogenous variation is directly useful for identifying parameters in a structural estimation. Ex ante, the answer is not obvious, as the number and strength of assumptions that go into establishing exogeneity can be at least as large as the number and strength of assumptions that go into a model.

Moreover, the answer to the question depends on the type of estimator one uses. For example, as Berry and Haile (2016) show, a broad set of estimators in industrial organization relies on plausibly exogenous variation. In particular, Berry, Levinsohn, and Pakes (1995), and Nevo (2000, 2001) consider estimation of models of imperfect competition between firms using data on market shares. Their estimation proceeds in two steps because identification of the parameters that govern the behavior of these firms requires as a first step identification of parameters that govern consumers' demand for each firm's product variety. However, recovering these utility parameters suffers from the usual simultaneity between supply and demand. Therefore, as Berry and Haile (2016) demonstrate using a nonparametric extension of these estimators, identification ultimately relies on instruments for product prices, which arguably affect the supply of each variety but not demand. Once the utility parameters have been estimated, the parameters governing the supply of each variety can then be identified from the firm's optimal behavior under imperfect competition. To summarize, in these two-step procedures, a verbal model is invoked implicitly to justify the instruments used to identify demand parameters, and a mathematical model is used explicitly to identify the parameters that determine competitive behavior.

Downloaded from https://academic.oup.com/rcfs/article-abstract/7/1/1/4590088
by Wuhan University Library user
on 02 March 2018

In contrast, the full-solution methods popular in corporate finance and macroeconomics, such as simulated method of moments, indirect inference, and simulated maximum likelihood, achieve identification quite differently by relying on relations between model predictions and model parameters, as described in Section 2.3.

Nonetheless, the question still remains whether one could achieve better identification if a source of plausibly exogenous variation were available. A necessary condition for any such variation to be useful is that the model contain a variable that corresponds closely to the exogenous variable in the data. This point is easiest to see in the context of an indirect inference estimator, which is based on the analogy principle that statistics calculated from a model should be set as close as possible as statistics calculated from data. Thus, these two sets of statistics need to be couched in terms of the same variables, so if one has an exogenous variable in the data, one needs to incorporate an analogous source of exogenous variation in the model.

However, this necessary condition that an exogenous variable be contained in both the model and the data is not sufficient for this exogenous variable to be useful for identification. There can be no general proof of this assertion, as the models typically estimated with simulation estimators have no closed form. However, it is straightforward to construct an instructive counterexample with the model that we have sketched above. To do so, we use a Monte Carlo simulation to compare two simulated minimum distance estimators of our model parameters. This setting allows us to distinguish exogenous from endogenous variation because in the laboratory of a Monte Carlo, the model is taken to be the truth.

For the first estimator, we construct moments (and functions of moments) using a measure of profits that is arguably observable in real data in the form of operating income, namely, $zk^{\alpha-1}$. This measure of profits contains both exogenous variation from z and endogenous variation from movement in the capital stock, k, just as operating income has both exogenous and endogenous components. For the second, we use a measure of profits, z, that is unobservable in real data, but observable and exogenous in the model.

For both estimators, we use the following set of moments to estimate the model, with this choice following closely the choices in Hennessy and Whited (2005). The first three are average profits, the variance of profits, and the slope coefficient from a regression of profits on lagged profits. These three quantities help identify α, σ, and ρ. Next we include average investment, which is strongly increasing in the depreciation rate, δ, and therefore useful for identification of this parameter. To identify the issuance cost parameter, λ, we include mean issuance and mean issuance incidence, that is, average $e(z,k,k')$, conditional on $e(z,k,k') < 0$, and the fraction of observations in which $e(z,k,k') < 0$. Finally, we include

Downloaded from https://academic.oup.com/rcfs/article-abstract/7/1/1/4590088
by Wuhan University Library user
on 02 March 2018

two additional moments, the variance of investment and the slope coefficient from regressing investment on profits. The variance of investment is useful for identification of α. If the firm has sharply decreasing returns technology, investment does not respond to profit shocks strongly and has a low variance. From Figure 1, the final moment has sharp, monotonic relations with several model parameters and should aid in their identification.

We next run two Monte Carlo simulations, following Bazdresch, Kahn, and Whited (forthcoming), to determine which estimator performs better in terms of parameter recovery in a finite sample. Each Monte Carlo is based on 1,000 simulated data sets in which the data are simulated from the model in Section 2. We consider a sample size of length 10 and cross-sectional width 500. This sample size is somewhat smaller than those typically available to corporate finance researchers, but we wish to be conservative. Finally, for both estimators, we use a clustered weight matrix, as this choice of weight matrix often results in better finite sample performance (Bazdresch, Kahn, and Whited forthcoming).

The results are in Table 1. For each parameter, we report the mean percentage bias over the 1,000 trials, as well as the root mean squared error (RMSE). The first important result is that all of the parameters have low bias and low RMSEs. The parameter with the highest bias and RMSE is λ, but even in this case, these figures are small. For example, a bias of –0.87%, which we see in Column (1) for the estimator using endogenous profits, can only be seen in the fourth decimal place in a parameter whose true value is 0.05. This result echoes similar results from Bazdresch, Kahn, and Whited (forthcoming), which conducts much more extensive Monte Carlo simulations of simulated minimum distance estimators. The intuition is simple. The statistics used in structural estimation, such as means, variances, and regression coefficients are themselves easy to estimate and thus are nearly unbiased in finite samples. Because the model parameters are functions of these simple statistics, they inherit this good performance. This last statement is true as long as the parameters are well identified, that is, as long as the mappings from the moments to the parameters are steep and monotonic.

The second and more important result is that using exogenous variation does not help a great deal in the estimation of the model parameters. Not surprisingly, using z as a measure of profits improves the performance of the estimator for the parameter ρ, with the RMSE falling by a third relative to the case in which we use $zk^{\alpha-1}$ as a measure of profits. This result makes sense inasmuch as ρ characterizes the serial correlation of z, and not the serial correlation of $zk^{\alpha-1}$. However, in both cases, estimator performance is excellent, and for all other parameters, the performance of the two estimators is extremely close.

Downloaded from https://academic.oup.com/rcfs/article-abstract/7/1/1/4590088
by Wuhan University Library user
on 02 March 2018

Table 1
Exogenous versus endogenous variation

	(1) Endogenous profits	(2) Exogenous profits
α (curvature)		
Mean percentage bias	0.1870	0.3521
Percentage RMSE	0.6855	0.8065
ρ (serial correlation)		
Mean percentage bias	0.1817	0.0196
Percentage RMSE	1.4873	1.0344
σ (standard deviation)		
Mean percentage bias	−0.1209	−0.2205
Percentage RMSE	0.7107	0.7206
δ (depreciation)		
Mean percentage bias	0.1158	0.2710
Percentage RMSE	1.0962	1.3812
λ (issuance costs)		
Mean percentage bias	−0.8707	−0.6758
Percentage RMSE	4.5302	5.0668

Indicated expectations and probabilities are estimates based on 1,000 Monte Carlo samples of size 5,000. The samples are generated from the model in Section 2. We consider two estimators. For both, we estimate model parameters using eight moments: the mean, variance, and serial correlation of profits; the mean and variance of investment; the mean and incidence of equity issuance; and the coefficient from regressing investment on profits. For the first, all profit moments are constructed using an endogenous measure of operating profits. For the second, all profit moments are constructed using a measure of profits that is by definition exogenous in the model. For each parameter, we report two statistics. Bias is expressed as a percentage of the true coefficient value. RMSE indicates root mean squared error and is also expressed as a percentage of the true coefficient.

This result makes sense given the result in Figure 1 that all model parameters are related to the sensitivity of investment to profits. Thus, this source of exogenous variation does not pertain to a specific parameter, just as many sources of exogenous variation in reduced-form studies do not pin down specific economic mechanisms that underlie estimated causal elasticities. Instead, identification comes from the structure of the model itself, which provides the correspondences between model predictions and parameters used to form the econometric objective function. This last point reinforces the general notion that identification requires assumptions, and that these assumptions can be either verbal or mathematical.

Of course, there might be certain circumstances, as in Bennedsen et al. (2007), in which one could find exogenous variation that pertains to a specific parameter. However, in that case, it would be better to run a regression than estimate that specific parameter as part of a structural estimation, with the caveat that the reduced-form approach is preferable as long as the verbal assumptions that tie the exogenous variation to a parameter are plausible.

Downloaded from https://academic.oup.com/rcfs/article-abstract/7/1/1/4590088
by Wuhan University Library user
on 02 March 2018

4. Conclusion

All our examples and our simple experiments with a dynamic model illustrate the importance of a model for identification. They point out that the model allows for the identification of an interesting economic quantity with a statistical parameter by advancing an internally consistent set of assumptions. A mathematical model is not essentially better or worse for this purpose than a verbal one. But careful advancement of a set of assumptions is difficult to accomplish verbally. Lawyers and researchers in the humanities practice for years to make their verbal arguments internally consistent. Economists are rarely so well prepared when they venture into a verbal model, but we have a great deal of experience with making mathematical arguments. So while mathematical models are not essentially better, they are often easier for economists to apply. Still, the theory is what allows us to identify structural parameters from statistical quantities.

All of our examples also illustrate that the question to be addressed comes before the model or the natural experiment. The model, with all of its assumptions, or the natural experiment, with its accompanying assumptions, is only a tool. It would be difficult to address the in-family CEO succession question with the estimation of a dynamic model, which would need to incorporate both product market conditions and family dynamics. In addition, as pointed out above, it would be impossible to assess the affects of agglomeration externalities with a natural experiment. More generally, there is no one approach that will be useful for answering all questions. But for whatever kind of question one asks, the key to identifying relevant parameters is to proceed guided by (either verbal or mathematical) theory and conscious of the necessary assumptions.

References

Andrews, I., M. Gentzkow, and J. M. Shapiro. 2017. Measuring the sensitivity of parameter estimates to sample statistics. *Quarterly Journal of Economics* 132:1553–1592.

Angrist, J. D., and V. Lavy. 1999. Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics* 114:533–75.

Angrist, J. D., and J.-S. Pischke. 2008. *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Ashenfelter, O. 1978. Estimating the effect of training programs on earnings. *Review of Economics and Statistics* 60:47–57.

Ashenfelter, O., and D. Card. 1985. Using the longitudinal structure of earnings to estimate the effect of training programs. *Review of Economics and Statistics* 67:648–60.

Bazdresch, S., R. J. Kahn, and T. M. Whited. Forthcoming. Estimating and testing dynamic corporate finance models. *Review of Financial Studies*.

Bennedsen, M., K. M. Nielsen, F. Pérez-González, and D. Wolfenzon. 2007. Inside the family firm: The role of families in succession decisions and performance. *Quarterly Journal of Economics* 122:647–91.

Downloaded from https://academic.oup.com/rcfs/article-abstract/7/1/1/4590088
by Wuhan University Library user
on 02 March 2018

Bernstein, S. 2015. Does going public affect innovation? *Journal of Finance* 70:1365–403.

Berry, S., and P. Haile. 2016. Identification in differentiated products markets. *Annual Review of Economics* 8:27–52.

Berry, S., J. Levinsohn, and A. Pakes. 1995. Automobile prices in market equilibrium. *Econometrica* 63:841–90.

Breza, E. 2012. *Peer effects and loan repayment: Evidence from the Krishna default crisis*. Manuscript, Harvard University.

Card, D., and A. B. Krueger. 1994. Minimum wages and employment: A case study of the fast food industry in New Jersey and Pennsylvania. *American Economic Review* 84:772–93.

Chernozhukov, V., H. Hong, and E. Tamer. 2007. Estimation and confidence regions for parameter sets in econometric models. *Econometrica* 75:1243–84.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan. 2011. How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *Quarterly Journal of Economics* 126:1593–660.

Davis, M. A., J. D. M. Fisher, and T. M. Whited. 2014. Macroeconomic implications of agglomeration. *Econometrica* 82:731–64.

Dube, A., T. W. Lester, and M. Reich. 2010. Minimum wage effects across state borders: Estimates using contiguous counties. *Review of Economics and Statistics* 92:945–64.

Glover, B., and O. Levine. Forthcoming. Idiosyncratic risk and the manager. *Journal of Financial Economics*.

Gomes, J. F. 2001. Financing investment. *American Economic Review* 91:1263–85.

Heckman, J., and R. Pinto. 2015. Causal analysis after Haavelmo. *Econometric Theory* 31:115–51.

Heckman, J. J., H. Ichimura, and P. E. Todd. 1997. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies* 64:605–54.

Hennessy, C. A., and T. M. Whited. 2005. Debt dynamics. *Journal of Finance* 60:1129–65.

———. 2007. How costly is external financing? Evidence from a structural estimation. *Journal of Finance* 62:1705–45.

Kahn, R. J., and T. M. Whited. 2016. Identification with models and exogenous data variation. *Foundations and Trends in Accounting* 10:361–75.

Koopmans, T. C. 1949. Identification problems in economic model construction. *Econometrica* 17:125–144.

Krueger, A. B. 1999. Experimental estimates of education production functions. *Quarterly Journal of Economics* 114:497–532.

Krueger, A. B., and D. M. Whitmore. 2001. The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR. *Economic Journal* 111:1–28.

Li, S., T. M. Whited, and Y. Wu. 2016. Collateral, taxes, and leverage. *Review of Financial Studies* 29:1453–500.

Marshall, A. 1890. *Principles of economics*. London: Macmillan and Company.

Moyen, N. 2004. Investment-cash flow sensitivities: Constrained versus unconstrained firms. *Journal of Finance* 59:2061–92.

Neumark, D., and W. Wascher. 1992. Employment effects of minimum and subminimum wages: Panel data on state minimum wage laws. *Industrial and Labor Relations Review* 46:55–81.

Downloaded from https://academic.oup.com/rcfs/article-abstract/7/1/1/4590088
by Wuhan University Library user
on 02 March 2018

Nevo, A. 2000. A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics and Management Strategy* 9:513–48.

———. 2001. Measuring market power in the ready-to-eat cereal industry. *Econometrica* 69:307–42.

Sorkin, I. 2015. Are there long-run effects of the minimum wage? *Review of Economic Dynamics* 18:306–33.

Staiger, D., and J. H. Stock. 1997. Instrumental variables regression with weak instruments. *Econometrica* 65:557–86.

Strebulaev, I. A., and T. M. Whited. 2012. Dynamic models and structural estimation in corporate finance. *Foundations and Trends in Finance* 6:1–163.

Downloaded from https://academic.oup.com/rcfs/article-abstract/7/1/1/4590088
by Wuhan University Library user
on 02 March 2018