Critical Finance Review, 2012, 2: 193–215

Model Before Measurement*

Christopher A. Hennessy

LBS, CEPR, and ECGI. Regent's Park, London NW1 4SA, United Kingdom; chennessy@london.edu

ABSTRACT

Empirical testing in corporate finance often proceeds from qualitative theory to hypothesis tests by way of verbal plausibility arguments or via static models that do not actually make predictions regarding the quantities measured in the data such as investment rates and leverage ratios. Examples are provided regarding how this methodology has resulted in faulty inference. It is proposed that the formulation of hypothesis tests should instead be informed by operational models that explicitly map the primitive theories to measured variables. Examples are provided for how a particular class of operational models, Dynamic Quantitative Models, can be used to: inform empirical tests; estimate parameters; extract information from observed response elasticities; and perform a broad range of quantitative policy experiments. Recent criticisms of such models are addressed, and future directions discussed.

1 Introduction

Empirical corporate finance has recently begun to speak the language of non-structural labor economics — almost exclusively. Often, the primary question that arises in an empirical seminar is "What's your identification strategy?" This is an important question, and worthy of attention. The

* I thank Joao Gomes, Terrence Hendershott and Joshua Rauh for detailed feedback. Hennessy acknowledges European Research Council grant funding.

ISSN 2164-5744; DOI 10.1561/104.000000010

purpose of this essay is to argue that there are other questions of equal importance that currently receive insufficient attention. Of course, a commonly voiced concern is that economics is becoming devoted to cleanly rejecting null hypotheses that nobody took seriously in the first place, or to cleanly delivering answers to questions that nobody had ever thought sufficiently interesting to pose before instruments were found to answer them. Perhaps the most forceful exponent of this view is Rust (2010), who worries that economics is devolving into an "elite infotainment industry." In support of his view, prior to the recent crises, one may have thought that gang finances, pop culture, and professional sports were of greater interest to economists than understanding the determinants of real investment, capital structure, and financial stability.

While Rust's concerns are certainly worth keeping in mind, to date his criticisms seem most applicable to the "general interest" economics journals. The best work in empirical corporate finance is still question-driven, not instrument-driven. Moreover, the recent trend in empirical corporate finance in thinking carefully about identification is a healthy development. It shows that researchers in the field are not settling on a fixed stock of tools. They are willing to import methods from other fields in order to make better inferences.

Recent trends notwithstanding, the cross-fertilization of empirical corporate finance is largely confined to non-structural labor, development, and public finance. As a result, the methodological tool-kit is artificially limited. In particular, the field is trapped in the past when it comes to thinking about dynamics, a topic that is central to asset pricing and macroeconomics. Hypothesis tests are still informed by static models despite the fact that we work with panel data. The economic distinction between stock and flow variables is generally viewed as "hair splitting", and the two are freely interchanged. Predictions about levels are prosaically transformed into predictions about differences, ratios, or differences in ratios. Claims are made about plausible degrees of leverage persistence based upon static models. The list goes on and on.

The neglect of dynamics is a symptom of a more general problem in empirical corporate finance. We seem to be racing towards an era of model-free empirical work. Of course, one usually sees the obligatory slide at seminars listing various models as motivation. Here, the revealed preference is for more abstract models, or classic papers. Ostensibly, this proves the empirical work is "serious" and "theory-driven." However, it is striking how often the cited papers are actually silent on the variables that are measured in the empirical tests. For example, empiricists commonly test predictions regarding leverage ratios when the underlying theory fails to make any clear prediction about the time-path of the capital stock. How does one predict a ratio without predicting its denominator?

The main point that I want to make in this paper is simple. The first question to be addressed in good empirical work is: "What does the theory actually predict here?" Despite the obvious importance of this question, it currently receives relatively little attention, perhaps being crowded out by intense and lengthy debates over whether exclusion restrictions are satisfied. In general, a verbal plausibility argument is viewed as sufficient to justify empirical tests. One may argue that there is no problem with this, and there are indeed many times when verbal reasoning leads to correct conclusions. However, I would argue that the accumulated stock of false falsifications and abandoned research lines suggests that we should move away from existing norms and insist upon a tighter link between theory and testing. This can be accomplished by mapping theories into *operational models.* As I describe below, an operational model is a model that delivers predictions about the variables that are used in empirical tests. *Dynamic Quantitative Models* (DQM) are an important example of the type of models that I have in mind. DQM are dynamic operational models that generate both qualitative and quantitative predictions.

Even when the underlying theory is clear about the sign of a predicted regression coefficient, there are still good arguments for placing a model before measurement. Using a model allows us to make an informed assessment of the plausibility of the *magnitude* of an empirical moment. Some of the most important puzzles in asset pricing are *quantitative*. Asset pricing models have evolved over time in order to address quantitative weaknesses of predecessors. This is one path to progress. In contrast, corporate finance has been reluctant to talk about magnitudes. This is unfortunate since magnitudes can be used to assess competing theories and identify ways in which theories need to be augmented at an operational level, if they are to better fit stylized facts. In addition, filtering data through a model allows one to estimate its underlying parameters. In turn, a parameterized model can serve as a powerful and general framework for policy analysis.

There is also a case for taking a model-based approach to instrumental variables (IV) work. Model building forces one to think carefully about the real-world data generating process. This can strengthen or weaken the case for satisfaction of the respective exclusion restrictions, by, for example, making key *a priori* assumptions explicit. Having a model in-hand influences the inferences we draw. For example, model building alerts one to the fact that treatment responses are parameter-contingent. Conversely, since treatment responses are parameter-contingent, estimated response magnitudes can be used to revise our beliefs about the magnitude of various parameters relating to firms' choice margins. Furthermore, the improved parameterized models can be used to think about a wider range of policy experiments without waiting for the experiments to actually occur in reality.

The remainder of this essay is as follows. Section 2 responds to Welch (2013) with the goal of clearing up misconceptions. Sections 3 discusses methodology. Section 4 discusses the problem of false falsification in empirical corporate finance. Section 5 describes the nature of statistical inference in the context of DQM, and then discusses how statistically parameterized DQM can be used for practical purposes. Section 6 discusses IV techniques and Section 7 offers concluding remarks.

2 A Response to Welch

My primary objection to Welch (2013) is that he debates an opponent of his own construction. Hennessy and Whited (HW) (2005) did not claim to solve the capital structure puzzle or assert primacy of the particular theory or model employed. Second, Welch ignores the fact that HW 2005 was intended as a *response* to a capital-structure-anomaly literature to which he had contributed, see for example Welch (2004). Third, and this is related, Welch attributes to us the very reduced-form regressions that we mimicked in order to critique. In doing so, he stands on its head our critique of the atheoretical reduced-form literature. Fourth, in his conclusion, Welch claims we constructed a model to strategically match one stylized fact. To the contrary, Welch has focused on one prediction in HW 2005 and has simply ignored all the others. Some stylized facts replicated by the model were considered falsifications of tradeoff theory. Other predictions cast doubt on the informativeness of the types of reduced-form regressions favored by Welch and others.

The following recounts what was actually stated in HW 2005. From the abstract: "there is no target leverage ratio, firms can be savers or heavily levered, leverage is path dependent, leverage is decreasing in lagged liquidity, and leverage varies negatively with external finance weighted average Q. Using estimates of structural parameters, we find the simulated model moments match data moments." We also point out that the tradeoff theory, as modeled, predicts a role for the financing gap and for real investment in predicting issuance activity, just as Myers (1984) assumes is the case in his pecking-order. Some may be tempted to ignore the implications of the models in HW 2005 and HW 2007. They suggest that there is no reason to expect leverage ratios to revert to a constant target as posited in Welch (2004). They cast doubt on granting the pecking-order a priority claim on the financing gap as a predictive variable in leverage regressions as, for example Fama and French (CFR, 2012), and numerous others have done. Finally, the respective moment-matching exercises show that Welch (2013) is incorrect to suggest tradeoff-based models, as a class, cannot match the frequency of issuance activity.

The data moments that are matched are: Average Investment/Assets, Variance of Investment/Assets, Average EBITDA/Assets, Average Net Debt/Assets, Average Equity Issuance/Assets, Frequency of Equity Issuance, Investment-Q Sensitivity, Debt-Q Sensitivity, Serial Correlation and Standard Deviation of Income/Assets. In HW 2007 we match an even broader sense of moments: Average Equity Issuance/Assets, Variance of Equity Issuance/Assets, Variance of Investment/Assets, Frequency of Equity Issuance, Average Payout Ratio, Variance of Distributions, Average Debt/Assets, Covariance of Investment and Equity Issuance, Covariance of Investment and Leverage, Serial Correlation and Standard Deviation of Income. Given all that we do quantitatively, I am puzzled that Welch characterizes our model as "not principally a quantitative model" and that he accuses us of attempting to match *one* stylized fact.

One need only read the first three paragraphs of HW 2005 to understand that we were certainly not proposing any reduced-form regressions. Rather, we were mimicking regressions in prior published papers in order to question the inferences that were drawn. The first paragraph of HW 2005 cites Miller (1977) and states: "This benchmark model has provided intuition and guidance for much of the empirical literature on corporate capital structure, which has uncovered several patterns in the data that are inconsistent with the static trade-off theory." The next paragraph goes on to list the patterns. First Graham (2000) is quoted: "Paradoxically, large, liquid, profitable firms with low expected distress costs use debt conservatively." This explains the emphasis that simulated firms can be savers in the abstract. Next, Myers (1993) is quoted: "The most telling evidence against the static trade-off theory is the strong inverse relationship between profitability and financial leverage". This explains the emphasis on the negative relationship between leverage and lagged liquidity in the abstract. Finally, there is a quote from Baker and Wurgler (2002): "The trade-off theory predicts that temporary fluctuations in the market to book ratio or any other variable should have temporary effects . . . capital structure is the cumulative outcome of attempts to time the equity market." This explains the abstract's emphasis on leverage persistence and the model's predicted negative relationship between leverage and lagged external finance weighted average Q. The third paragraph begins: "This paper shows that a dynamic trade-off model can explain *these* [my italics] stylized facts."

The narrow claim made at the conclusion of the paper is that the three anomalies are not anomalies at all once tradeoff theory is made operational using a dynamic model. Our conclusion stated: "In recent years, corporate finance economists have uncovered a variety of phenomena that appear anomalous in light of the traditional static tradeoff theory. Our paper questions whether these phenomena are indeed anomalies." Based on demonstrating this narrow claim, we did hope to make a more general point. By illustrating the fallibility of atheoretical reduced-form work, the paper was intended to influence methodology in empirical corporate finance. The focus on tradeoff theory was coincidental, and driven by concerns about model tractability and transparency. In particular, we had hoped that the relative simplicity of a tax-based model would allow us to cleanly deliver this message: If the best and brightest researchers have repeatedly incorrectly guessed the empirical implications of what is, perhaps, the simplest theory of financial structure, what confidence do we have in their ability to correctly guess the empirical implications of more complex theories?[1]

The disagreement I have with Welch is not primarily along the lines of the empirical merits of alternative theories. Rather, the disagreement centers on methodology. The primary objective of many DQM projects over the past decade has been to use simulated models as laboratories illustrating the inherent unreliability of two categories of empirical tests. First, null hypotheses based on verbal plausibility arguments instead of operational models that speak directly to the empirical moment considered. Second,

[1] Similar arguments have been made in response to the asset pricing anomaly literature.

alleged theory-driven tests where the model serving as the motivation for the test does not actually generate predictions about the empirical moment that is considered.

Welch's informal attempts at testing the predictions of dynamic tradeoff theory only serve to reinforce our original argument against atheoretical reduced-form work. Welch replicates just one of the mimicking regressions in HW 2005, leverage on *cash flow* (we also use EBITDA/Capital as a measure of liquidity based on the prior reduced-form literature), and obtains a negative coefficient, consistent with the model. Inexplicably, he then performs a number of regressions of his own choosing in which he replaces cash flow with the *cash stock* as an explanatory variable.[2] Cash flow and cash stock are different variables. The former is an exogenous flow while the latter is an endogenously chosen stock, so they cannot be freely interchanged, and there is no reason to expect them to have similar regression coefficients. Despite this fact, based on his leverage-cash stock regressions Welch (2010) originally argued: "Even in sample, the data strongly reject these models." To the contrary, existing models suggest that Welch has incorrectly specified the null hypothesis with respect to the relationship between leverage and the cash stock that would be predicted under tradeoff theory. In particular, HW 2005 shows that tradeoff theory predicts leverage is decreasing in lagged cash flow. Intuitively, simulated firms substitute cash windfalls for external debt. In contrast, Gamba and Triantis (2010) extend the model of HW 2005 to allow for simultaneous holdings of cash stocks by endogenously levered firms in a tradeoff theoretic setting. This is a model that actually speaks to the regressions that are run by Welch. They show that cash stocks can act as a *complement* for leverage. Intuitively, a firm that simultaneously leverages and holds a cash stock can use reductions in the cash stock in future periods as a means of increasing net leverage while avoiding debt issuance costs.

3 Some Thoughts on Methodological Norms in Empirical Corporate Finance

Let me clarify what I mean by methodology. To fix ideas, let us assume that the question of interest is: How do firms determine their financial

2 In justifying the substitution, Welch cites Bolton *et al.* (2010), a model with no taxes in it.

structures? We have a non-exhaustive array of competing theories. For example, we might be interested in assessing the relative merits of the following theories. (A) financial structure reflects firms optimally trading off tax benefits of debt against costs of financial distress; (B) financial structure is chosen optimally in response to moral hazard; (C) financial structure is chosen optimally in light of informational asymmetries between insiders and rational outsiders; and (D) financial structure is chosen optimally by insiders to exploit behavioral biases of outsiders. We then have an array of models in which the respective theories are made operational. For example, HW 2005 and HW 2007 make Theory A operational. Next, with the theory and operational model in-hand, one proceeds to formulating null hypotheses, for example, "According to Theory A, the regression coefficient on variable Z should be positive." After estimating the regression coefficient or some other moment in real-world data, we reach some conclusion. Continuing with the preceding example, the falsificationist might say: "The coefficient on variable Z is positive so we have simply failed to falsify Theory A, just as Theory D remains unfalsified. Let's look for some bolder predictions made by the respective theories to see if they can be falsified." The positivist might be more optimistic saying: "The fact that the coefficient on variable Z is positive constitutes favorable empirical evidence for Theory A."

This sequence of steps: (1) theory; (2) operational model; (3) null hypotheses; and (4) conclusions are the components of a sound methodology. The primary methodological objection raised in HW 2005, HW 2007 and Hennessy *et al.* (2010) (HLM) concerns Step 2. In my opinion, Step 2, making the underlying theory *operational*, is often bypassed altogether in empirical work, with a verbal plausibility argument put in its place. I call this style of empirical research atheoretical because the null hypothesis is actually divorced from the theory.

In order to make a theory operational, one should start with an explicit statement of the objective function, technology, contracting space, constraints, sources of uncertainty, information structure, and the equilibrium concept. From there, one uses the operational model to generate predictions about the relationship between endogenous variables and observables. For example, the leverage ratio can be related to an observable variable, such as the market-to-book ratio, or parameters, such as bankruptcy costs and tax rates. Ideally one can perform comparative statics analytically. In other cases, numerical simulation may be necessary. In either case, the data generating process in the model laboratory should mimic the data generating process that provides the context for the empirical test. For many empirical tests, justifying the null hypothesis will be easy, which is fine. The point is to force all empirical work to at least clear this minimal internal consistency hurdle to weed out suspect tests and suspect lines of research.

Welch (2013) states: “In the absence of better empirical support, the Hennessy and Whited (2005) model should not be the primary lens through which capital structure is viewed.” This comment reveals a fundamental misunderstanding. I do not think the HW 2005 or HW 2007 models should be the primary lens through which capital structure is viewed. However, I do think that such models, featuring endogenous debt, investment, and retentions merit being the primary lens through which we view *empirical tests* of Theory A (tradeoff theory). Similarly, I do not think that the model of DeMarzo and Fishman (2007) should be the primary lens through which capital structure is viewed. However, I do think such a model merits being a primary lens through which we view empirical tests of Theory B (moral hazard). I do not think the HLM 2010 model should be the primary lens through which capital structure is viewed. However, I do think such a model merits being a primary lens through which we view empirical tests of Theory C (asymmetric information).

The reader will notice that I exclude Theory D from my list. This is deliberate. In my view, it is startling how many of our leading theories of financial structure have been “tested” without stating any assumptions about how they are made operational. The market-timing theory of Baker and Wurgler (2002) is a primary example. Here all assumptions are implicit, with most endogenous quantities treated as free-variables. They do not specify an objective function or constraints. Are we to believe that market-timing managers can issue infinite quantities of overpriced equity without investors reacting? Of course not. Clearly, the authors must be making some implicit assumptions about boundedly rational Bayesian updating here, but these are never stated. Are we to believe that only the HW 2005 model features the “deep structure” of a profit function, as Welch (2013) argues? Hardly. Every truly operational model of financial structure needs to make assumptions about how physical capital maps to profits, and how firms choose their investment. To see this, note that capital is the denominator in most of the ratios we use in the literature on financial structure and real investment.

If the capital stock changes, so too does the leverage ratio, Q ratio, market-to-book ratio, profitability, etc.

The pecking-order "theory" is especially problematic methodologically since it is actually a convex combination of Theory C and a set of null hypotheses. The pecking-order theory is actually a set of hypotheses that are not derived from an explicit statement of the objective function and constraints. Rather, Myers (1984) actually used observed financing policies in leapfrogging from Step 1 to Step 3. This is a rather troubling way to ensure a theory delivers predictions fitting the stylized facts. Myers (1984) and Myers and Majluf (1984) also shift between separating and pooling equilibria without mentioning either. But the two equilibrium concepts have very different empirical implications.

Nachman and Noe (1994) place the pecking-order theory on a more solid foundation, showing that under technical conditions firms will finance with debt in a pooling equilibrium. Their model suffers from a number of weaknesses in terms of matching the data. First, they assume an exogenously fixed level of real investment. In reality, capital expenditures are variable, can act as a signal, and can play a critical role in explaining financing activity. Second, their static model is necessarily silent on retention policy. Third, their model cannot explain why some firms use equity as their first source of funding. Finally, if markets only implemented pooling equilibria, we would not observe announcement effects surrounding changes in financing or investment. HLM 2010 develop an operational model of Theory C which allows for separating and pooling equilibria in order to address these empirical shortcomings. Still, I fear that empiricists will continue assuming Theory C = pecking-order for some time. See, for example, Fama and French (2012).

To be clear, I am not arguing for the unimportance of the market-timing theory proposed by Baker and Wurgler (2002) and the adverse selection problem flagged by Myers and Majluf (1984). To the contrary, one may reasonably hold a prior belief that these theories are more compelling than others. However, if this is the case, the theories deserve a logically coherent exposition that is clear about underlying technologies, constraints, and the equilibrium concept. This should be done before one proclaims empirical successes or failures. In other words, before an empiricist documents a puzzle, anomaly or success, they should have an informed prior regarding what the underlying theory actually says about the chosen regression or empirical moment. In order to do this, the theory has to be made operational (Step 2).

Welch (2013) does not view the absence of an operational model, as a barrier to good empirical work. Instead, he cites as exemplary empirical work Shyam-Sunder and Myers (1999) who "tested" pecking-order theory against the tax-distress-friction model and "promptly found it to perform as well." He also favorably cites Huang and Ritter (2009) and Fama and French (2012), who manage to "test" the static tradeoff model, using panel data, against the pecking-order model (sic) and market timing model (sic). Frankly, in the absence of operational models that speak to the panel data utilized, I have little idea of the meaning of the regression coefficients that are reported in these papers.

4 The Problem of False Falsification

The falsificationist methodology is ruthlessly efficient. After all, once a theory has been falsified, we can eliminate it from consideration and move on to others. But this very ruthlessness implies that one must take special care that empirical tests do not falsify incorrectly. I refer to the problem of incorrect rejection of a true theory as *false falsification*. I treat it as distinct from a Type I error since many view Type I errors in a narrow statistical sense. Large standard errors and unlucky draws are however only one source of false falsification. Perhaps a more common source of false falsification in empirical corporate finance is the failure to base null hypotheses on an operational model that actually speaks to the sign and magnitude of the moment under consideration.

The insistence on maintaining a tight map from theory to model to measurement is a reaction to the casual way in which the profession, theorists as well as empiricists, implicitly advocates moving from theory to hypothesis tests. Consider the following simple example. Suppose an elegant static model predicts that private information causes the capital stock (K) and firm value (V) to fall below first-best. Now suppose that an empiricist wants to see whether private information is operative in reality. Finally, suppose that the empiricist has hand-collected an ideal panel data set in which half of the firms were randomly and publicly assigned the problem. The data set contains information on capital stocks, annual investment, and market capitalization for each firm over ten years. The empiricist would surely use V/K and I/K in her empirical tests. After all, "we always work with Q ratios and investment rates." But what does the original static theory actually tell us about how average Q ratios and investment rates should differ between the two halves of the sample? It is tempting to make the leap from "lower V" to "lower V/K" and from "lower K" to "lower I/K" but it is clear that lower K may actually lead to an increase in Q ratios and investment rates. Indeed, HLM 2010 confirm that this is the case in a calibrated simulation of this scenario. Thus, the static theoretical model is quite misleading about the dependent variables here.

Let us however continue with the preceding example. Under current norms, it is a fair guess that many empiricists would proceed to "testing the theory" with the null hypothesis that "the theory predicts lower Q ratios and investment." Furthermore, the original theorist sitting in the seminar room would probably not even catch the shift. The point here is that the lack of precision in empirical corporate finance is so ingrained that neither theorist nor empiricist would realize that any leap was being made. Moreover, even when all parties, including the original theorist, participate in determining empirical implications, the chasm between the original model and observables is often too wide to bridge.

There is much more work to doing good theory-driven empirical research in corporate finance than reading the original papers. The majority of our most compelling theories were exposited using stylized static models. Such models are ideal for providing intuition regarding causal mechanisms. They also strip away analytical complexity allowing one to isolate key assumptions regarding what is observable, verifiable, the space of contracts, the feasibility of renegotiation, etc. However, the original stylized models are often ill-suited for making the underlying theory operational in the sense of speaking to the data. This is especially the case for static models. After all, when it comes to empirical testing in empirical corporate finance, we are most commonly working with ratios involving stock and flow variables such as Value/Capital, Investment/Capital, Debt/Capital, Cash Stock/Capital, EBITDA/Capital, and so on.

Now recall that the problem of false falsification arises when we falsify a true theory. But how do we know we have falsified a true theory when we do not actually know the truth? Here is where DQM can play an important role. DQM can serve as an ideal laboratory for testing our empirical tests. Constructing a DQM and simulating it allows us to know the true data generating process in the laboratory. By first conducting planned empirical tests on simulated data, we can consider the extent to which various empirical tests are vulnerable to false falsification. This approach is used by Gomes (2001), Cooper and Ejarque (2003), Alti (2003), Moyen (2004), and HW 2007 in the context of the empirical literature on real investment, and by HW 2005, Strebulaev (2007) and HLM (2010) in the context of the empirical literature on leverage ratios.

I should note here that while following my prescription mitigates the problem of false falsification, it does not eliminate it. After all, it is an operational model of a theory that is used to generate predictions, not "the theory" per se. An operational model may generate a prediction that is rejected in the real-world data, but that only allows us to reject that particular operational model, not the underlying theory. A given theory can be made operational in a number of ways and this can influence empirical implications. Therefore, if the goal is to attempt a falsification of a given theory, one should try to identify ineluctable predictions of the underlying theory that will necessarily be shared by all operational versions of it. Let me give an example that illustrates the prior argument, and that also serves to illustrate why I think that clarity is preserved if we make a distinction between Theory (Step 1) and Operational Model (Step 2). In the interest of tractability, Strebulaev (2007) assumes that the firm cannot issue debt incrementally. Rather, the firm must retire its old debt before issuing a new level of debt. Based on the infrequency of financing activity implied by Strebulaev's operational model, Welch (2013) argues: "The data suggest that *the* [my italics] tax-distress-friction model cannot possibly explain the empirical non-adjustment evidence in corporate leverage ratios that motivated it." I would argue that here Welch is incorrect in using "the" as this implies that there is only one operational model of the tax-distress-frictions theory and that this one model cannot explain the frequency of issuance activity. The necessary conclusion is that the *theory* must be rejected. However, all that has really been demonstrated here is that a particular operational model has been rejected. Other operational models, for example, those in HW 2005 and HW 2007, generate frequent issuance activity. So rejecting the theory on this basis would constitute false falsification.

Part of the appeal of mapping theories to DQM prior to empirical testing is that doing so forces one to be precise about variable definitions, in the model and in the real-world data. For example, informal treatments of the financing decision are usually deliberately imprecise about the definition of leverage. In contrast, in working with a DQM one can work with either book or market leverage, and make distinctions between their behavior. Another important feature of DQM is that there is a clear mapping between the objective function and controls. Managers are not assumed to optimize a leverage ratio. Instead, they choose real and financial control policies to maximize firm value. One can then compute an implied book or market leverage ratio at each point in time.

There are also times when being explicit about a problem, up to the level of the driving process for profits, can help one make sense of empirical puzzles. Let me give an example from HLM 2010. In the model, firms signal positive information by financing with debt and choosing a high level of capital while those with negative information are unlevered. Yet, in the simulated data there is a negligible announcement effect surrounding debt issuance. This model-based observation resembles the real-world observation of negligible debt issuance announcement effects as documented by Eckbo (1986). In the model, the absence of an announcement effect is puzzling while in the real-world data the absence is only puzzling if one believes that debt signaling is operative. We soon recognized that the simulated puzzle was driven by persistence in private information. A firm with positive information in one period tends to have positive information in the next period as well, and will tend to have high leverage in both periods. So, there will not be much of an abnormal return when this firm announces debt in the later period. However, if a firm switches from being unlevered to levered, or vice-versa, one sees a large abnormal return in simulated data. In fact, this simulated observation is consistent with the evidence in Nandy *et al.* (2008).

There are also times when an explicit treatment of a problem will make it clear that a given line of attack is futile. Let me give an example based on HW 2007. Part of the goal of the investment-cash flow research that was initiated by Fazzari *et al.* (1988) was to draw inferences about the magnitude of financing frictions from the coefficient on cash flow in an investment regression. Once one writes down the full problem parametrically it becomes obvious that this is a hopeless exercise. After all, there are numerous parameters describing financing frictions: issuance costs, collateral constraints, and bankruptcy costs, etc. How could one possibly hope to infer all these parameters from a single coefficient?

As a related example, consider that, following Fazzari *et al.* (1988), there was a cottage industry in financial constraint indexes. However, HW 2007 show that in simulated data, many of the indexes actually decrease when financing cost parameters are increased. The reason is simple. When financing costs increase, the simulated firm tends to conserve debt capacity, while most constraint indexes treat low leverage as indicative of being less constrained. HW 2007 thus argues that the constraint indexes are best viewed as proxies for limits on future funding capabilities as opposed to costs of external funds. They are two distinct economic concepts and are likely to move in opposite directions given optimizing behavior. Again, working through an operational model provides clarity where informal verbal treatments do not.

5 Structural Inference

Returning to the essay by Welch (2013), it is apparent that he misunderstands the nature of the statistical inference problems tackled in HW 2005 and HW 2007. When Welch uses the term inference, he has in mind a notion of inferring the true theory determining financial structure. Our goals were more modest.

The goal of structural inference is to best estimate of the magnitude of deep parameters. The goal is not to estimate elasticities or responses to treatment effects. However, once a model is parameterized, it can be used to estimate these quantities via simulated policy experiments. Estimating deep structural parameters is useful for a number of reasons. First, some of the parameters have a natural economic interpretation and are of direct interest, such as costs of financial distress. The second benefit of this inference procedure is that it allows one to iterate toward better operational models. For example, we have argued that large issuance cost and large bankruptcy cost estimates are suggestive of indirect costs not captured by our models. Finally, estimating structural parameters allows one to perform a broad range of quantitative policy experiments.

There are a number of techniques available for structural inference. In some cases one can express the likelihood function analytically in terms of model parameters so that maximum likelihood may be used. When this is intractable, simulated maximum likelihood can be employed. In contrast, a method of moments estimator chooses parameters to minimize some measure of the distance between a vector of model-implied moments and real-world moments. Simulated method of moments (SMM) chooses parameters to minimize the distance between real data moments and the moments generated in simulated model data. A near-relative of SMM is indirect inference, where an auxiliary model, such as a regression of investment rates on Tobin's Q and cash-flow, is fit to both real and simulated data, providing additional features to match.

There is some tension between statistical and other concerns in applying the method of moments. For example, tests on overidentifying restrictions allow one to test the null hypothesis that the specified model is the "true model." I suspect that most of us view such tests as leading to Type II errors with probability one since we will never find the one true model that captures the real-world data-generating process. The real goal is not to find the one true model. A more realistic goal is to improve upon earlier theories and earlier models. This brings up a second, more technical, point. Most researchers who use these methods choose their moment weighting matrix to minimize the asymptotic variance of their estimator. I think this is a mistake in the context of quantitative corporate finance. At this early stage in the field, the priority should be on reaching consensus regarding the most important moments to match. Once this has been established, the alternative models can be assessed in terms of their average cross-moment percentage error.

One frequently voiced criticism of SMM is that moments are chosen strategically in order to make the models look good. In reality, moment selection is driven by two concerns. First, the HW 2005 and HW 2007 papers selected ratios that any good model should fit. Admittedly, unlike asset pricing, we have yet to reach a consensus on which moments are key. Second, in using SMM, moment selection is driven by the requirement that the respective model moments are sensitive to changes in the underlying parameters, in the sense of the Jacobian determinant being nonzero. In fact, this is the basis for the identification of parameters in the context of SMM, not an orthogonal error condition. From this perspective, the ideal moments to select are those having the potential to make the simulated model look as bad as possible, and not those moments the model will always hit.

Let me close this section by noting that I have always viewed the moment matching exercises in HW 2005 and HW 2007 as a first step; laying down a gauntlet, so to speak. We were hoping to see new and improved dynamic models being taken to the data, beating us at our own game. This is now beginning to happen. For example, Nikolov and Schmid (2012) consider a variant of the hidden action/moral hazard model of DeMarzo and Fishman (2007). Morellec *et al.* (2013) formulate an operational model where the manager can divert free cash flow. There is still however ample work to do.

6 Instrumental Variables and Natural Experiments

Welch (2013) posits that there are a wide range of interesting questions that can be settled using IV techniques since valid instruments are in abundance. I would argue that valid instruments are scarce in supply. The very lengths that empiricists go to in order to find valid instruments testifies to their scarcity. At times, some lose sight of the reason why they started searching for the instrument in the first place. In the worst case scenario, an instrument is found in order to answer a question that nobody finds interesting except aficionado's of clever identification. Related to this point about the cart being placed before the horse, Heckman (1997) shows that when there is heterogeneity, the choice of instrument may determine the probability limit of an IV estimator. That is, the instrument is dictating the quantity estimated. Deaton (2010) argues that "this goes beyond the old story of looking for an object where the light is strong enough to see; rather, we have at least some control over the light but choose to let it fall where it may and then proclaim that whatever it illuminates is what we were looking for all along." In contrast, in a structural estimation, one first specifies the parameter to be estimated and then looks for moments that are informative about it. For example, the variance of investment is informative about parameters of the capital adjustment cost function.

As another example of the scarcity of valid instruments, Deaton (2010) argues that in many instances the IV literature has presumed a variable to be a valid instrument simply because it is externally determined. However, closer inspection may reveal that the external variable fails to satisfy the exclusion restrictions as it affects the dependent variable through channels other than the regressor of interest. As examples he cites the literature on aid and growth using *Egypt* and *francophone* as instruments when it is clear that being Egypt or francophone affects growth through channels other than the quantity of foreign aid. Going forward, it will be important to ensure the distinction between externality and exogeneity is maintained, especially given that the prior use of an instrument can be taken as a form of validation.

As another example of the scarcity of valid instruments, consider the suggestion by Welch (2013) that we look to tax reforms as natural experiments to determine how corporate behavior is influenced by taxation. A basic problem here is that tax code changes are never clean. For example, in discussing the Tax Reform Act of 1986 (TRA86) Auerbach and Slemrod (1997) argue: "One disadvantage of TRA86 as an event study is its extensiveness; hundreds of provisions concerning the definition of the tax base were changed, introduced or eliminated, not to mention the change in the rate structure applying to both the corporate and individual bases." There is another, more fundamental issue here. Suppose TRA86 featured just one change, a reduction in the corporate income tax rate. Even here, we do not have a randomized experiment since the rate change was not chosen at random, but was perceived as being optimal given the technologies and corporate structures prevailing at the time. Given these pitfalls, it seems to me that a structural approach, with policy experiments performed on a parameterized model, deserves serious consideration as an alternative means of assessing the consequences of corporate tax reform. This line of research was pursued by Ballard *et al.* (1985) but abandoned in light of the fact that public finance economists did not then have at their disposal operational models delivering interior optimal financial structures.

Let us set aside these issues and assume that Joshua Angrist is elected Treasury Secretary, with his first act being to randomly choose a new corporate income tax rate. Would that be enough to settle things? Of course, a first issue to confront is that the very willingness to voluntarily randomize suggests the perceived costs of doing so are sufficiently low.[3] Even when setting this issue aside, problems remain. In particular, I would argue that one still needs an operational model to know what the underlying theory predicts. Furthermore, filtering observed responses through the prism of an operational model would allow Secretary Angrist to extract information that is relevant to assessing future policy changes. He cannot continually randomize features of the tax code one by one. Eventually, he'll need an operational model from which he can feel confident in making generalizations. I will elaborate on these points below.

In struggling to reach any definitive conclusion on the effects of TRA86 on corporate behavior, Auerbach and Slemrod (1997) state: "But the difficulties of knowing how to measure leverage and how to control for

[3] See Kaplan *et al.* (2013) for such a voluntary randomization involving stock lending.

unrelated factors make the interpretation of aggregate evidence difficult." With an operational model in-hand, there is no correct or incorrect measure of leverage. And why should there be one correct measure of leverage? Once a theory has been made operational, it will deliver predictions regarding any measure of leverage (debt-to-market, debt-to-book, debt-to-equity, coverage ratios, etc), taking into account endogenous changes in capital stocks, debt stocks, and the prices of debt and equity. An informal verbal treatment would struggle to make correct predictions given all the margins of adjustment. Another example is provided by Gordon and MacKie-Mason (1991) who perform a careful analysis of TRA86, finding that the Miller corporate tax shield value increased from 19.9% to 22.4%. They argue that "the change in debt/value ratios has been substantially smaller than expected." However, the basis for measuring the expected increase in leverage was a prior empirical model, as the authors abandoned an attempt based upon a micro-founded weighing of tax benefits against bankruptcy costs. To be clear, I am not arguing against using IV and quasi-natural experiments. However, I am arguing that having an operational model in-hand can help determine what the theory actually predicts regarding responses. For example, given the relatively small increase in the effective tax shield value estimated by Gordon and MacKie-Mason (1991), perhaps one should expect roughly zero response given plausible magnitudes of issuance and capital adjustment costs. But it would be incorrect to infer that corporations would fail to respond to a large change in the tax shield value.

To continue with the preceding example, instead suppose that leverage ratios increased a great deal in response to an increase in the value of debt tax shields. Should we stop here? I would argue that there is much more information to be extracted from the scarce resource of tax rate changes. An operational model would tell us which margins we should expect to see affected: debt levels, retentions, real capital stocks, equity prices, etc. By filtering each real-world response margin through an operational model, one can draw inferences about the underlying parameters that influence those margins, such as issuance costs, distress costs, and capital adjustment costs. In turn, the newly-parameterized model can be used to think about policies that have not yet been implemented. In sum, leverage elasticities are not a sufficient-statistic although the public finance literature has treated them as such. Instead, we should work back to the level of deep parameters if we want to extract maximum information and if we want to be able to generalize to a broader range of policy regimes.

The beauty of the SMM approach is that there is no need to wait for tax rate changes to draw inferences about deep structural parameters, although such changes add to the stock of moments that can be used to draw such inferences. And since SMM forces one to work back to the level of deep parameters, one can conduct a broad range of policy experiments. This generality strikes me as a real advantage over policy-specific natural experiments. For example, observed responses to an increase in the investment tax credit should be informative about the likely effect of acceleration of depreciation allowances or a decrease in the corporate income tax rate. With an operational model in-hand, one can map conclusions from one policy change to another in an informed way. In contrast, a purely empirical approach forces us to test policies one by one. Not only is this approach practically infeasible, it also unduly constrains our ability to generalize.

7 Concluding Remarks

In this essay, I have argued that operational models should be more central to empirical work in corporate finance. As evidence, I have discussed prominent lines of research where the chosen regressions fail to settle anything. Many of these lines of research have been discredited. Despite this fact, hypotheses about leverage ratios, investment rates, Q ratios, and others, are still tested despite the fact that the underlying theory is often silent about the ratios considered. Just as we have ended the presumption of orthogonal errors, we should end the presumption that a verbally plausible argument justifies a null hypothesis. Null hypotheses should be justified formally. The recent trend towards identification is a good thing but will not cure the problem of false-falsification that I have flagged. Little is learned from a misinterpreted regression coefficient, even if clever identification renders it consistent.

I have also argued that even when the underlying theory's qualitative prediction is clear, it is still worth considering mapping the theory to an operational model. After all, the magnitude of empirical moments and regression coefficients are another means of assessing theories. If we are to have an informed null regarding magnitudes, we need to work through operational models. Furthermore, having an operational model in-hand allows one to extract more information from an observed elasticity or treatment response. Random assignment and good instruments are sufficiently rare for them to go to waste. At core, observed elasticities are functions of deep parameters. We should use observed elasticities to update our estimates of parameters. In turn, with updated parameter estimates in-hand, we can consider a broader range of policy experiments.

There is another reason I favor a move to operational models, specifically in the form of Dynamic Quantitative Models. By their nature, DQM generate predictions about a broad range of choice margins. In contrast, I fear that we are currently on a research path leading to nowhere where there is a theory for each stylized fact. "Leverage is chosen on the basis of a desired bond rating." "Dividends are chosen in order to cater to investor demand for them." "Investment is chosen according to managerial empire building motives." "Firms keep war chests anticipating product market competition." The list goes on and on. To be clear, I think it is worthwhile documenting such concerns and considerations (rigorously), but it is also high time for us to give our leading theories a logically coherent exposition and assess their ability to fit a broad range of stylized facts on both the real side as well as the financing side.

One can look to asset pricing to see how such a quantitative approach leads to progress. For example, Mehra and Prescott (1985) found that under isoelastic preferences, matching the observed equity premium demands an implausibly high coefficient of relative risk aversion. This shortcoming led to General Expected Utility, which allows for de-linkage of risk aversion and intertemporal substitution. The equity premium puzzle also led to models featuring internal and external habit formation. Moment-driven iteration towards better operational models continues to this day in asset pricing. This may not be the only path to progress in corporate finance but it is certainly a path that is woefully underdeveloped.

References

Alti, Aydogan. 2003. "How Sensitive is Investment to Cash Flow when Financing is Frictionless?" *Journal of Finance*, 707–722.

Auerbach, Alan J. and Joel Slemrod. 1997. "The Economic Effects of the Tax Reform Act of 1986." *Journal of Economic Literature* 35: 589–632.

Baker, Malcolm and Jeffrey Wurgler. 2002. "Market Timing and Capital Structure." *Journal of Finance* 57: 1–32.

Ballard, Charles, Don Fullerton, John Shoven, and John Whalley. 1985. *A General Equilibrium Model for Tax Policy Evaluation*. National Bureau of Economic Research Monograph.

Bolton, Patrick, Hui Chen, and Neng Wang. 2010. "A Unified Theory of Tobin's Q, Corporate Investment, Financing, and Risk Management." *Journal of Finance*.

Cooper, Russell and Joao Ejarque. 2003. "Financial Frictions and Investment: A Requiem in Q." *Journal of Economic Dynamics* 6: 710–728.

Deaton, Angus. 2010. "Instruments, Randomization, and Learning About Development." *Journal of Economic Literature* 48: 424–455.

DeMarzo, Peter and Mike Fishman. 2007. "Agency and Optimal Investment Dynamics." *Review of Financial Studies*, 151–188.

Eckbo, Espen. 1986. "Valuation Effects of Corporate Debt Offerings." *Journal of Financial Economics*, 115–151.

Fama, Eugene F. and Kenneth R. French. 2012. "Capital Structure Choices." *Critical Finance Review* 1: 59–101.

Fazzari, Steven, R. Glenn Hubbard, and Bruce Petersen. 1988. "Financing Constraints and Corporate Investment." *Brookings Papers on Economic Activity* 1: 144–195.

Gamba, Andrea and Alex Triantis. 2010. "The Value of Financial Flexibility." *Journal of Finance* 63: 2263–2296.

Gomes, Joao. 2001. "Financing Investment." *American Economic Review* 91: 1263–1285.

Gordon, Roger and Jeffrey MacKie-Mason. 1991. "Effects of the Tax Reform Act of 1986 on Corporate Financial Policy and Organization Form." in *Do Taxes Matter? Evidence from the Tax Reform Act of 1986*, Joel Slemrod ed., MIT Press.

Graham, John. 2000. "How Big Are the Tax Benefits of Debt?" *Journal of Finance* 55: 1901–1941.

Heckman, James. 1997. "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *Journal of Human Resources* 32: 441–462.

Hennessy, Christopher A., Dimitri Livdan, and Bruno Miranda. 2010. "Repeated Signaling and Firm Dynamics." *Review of Financial Studies*.

Hennessy, Christopher A. and Toni M. Whited. 2005. "Debt Dynamics." *Journal of Finance* 60: 1129–1165.

Hennessy, Christopher A. and Toni M. Whited. 2007. "How Costly is External Financing? Evidence from a Structural Estimation." *Journal of Finance* 62: 1705–1745.

Huang, Rongbing and Jay Ritter. 2009. "Testing Theories of Capital Structure and Estimating the Speed of Adjustment." *Journal of Financial and Quantitative Analysis* 44: 237–271.

Kaplan, Steve, Tobias Moskowitz, and Berk Sensoy. 2013. "The Effect of Stock Lending on Security Prices: An Experiment." forthcoming in *Journal of Finance*.

Mehra, Rajneesh and Edward Prescott. 1985. "The Equity Premium: A Puzzle." *Journal of Monetary Economics* 15: 145–161.

Morellec, Erwan, Boris Nikolov, and Norman Schuerhoff. 2012. "Corporate Governance and Capital Structure Dynamics." *Journal of Finance*, 803–848.

Moyen, Nathalie. 2004. "Investment-Cash Flow Sensitivities: Constrained versus Unconstrained Firms." *Journal of Finance* 69: 2061–2092.

Myers, Stewart. 1984. "The Capital Structure Puzzle." *Journal of Finance* 39: 575–592.

Myers, Stewart. 1993. "Still Searching for the Optimal Capital Structure." *Journal of Applied Corporate Finance* 6: 4–14.

Myers, Steward and Nicholas Majluf. 1984. "Corporate Financing and Investment Decisions when Firms have Information that Investors do not Have." *Journal of Financial Economics* 13: 187–221.

Nandy, D., M. Kamstra, and P. Shao. 2008. "Do Financing Expectations Affect Announcement and Long-Run Stock Performance." working paper, York University.

Nikolov, Boris and Lukas Schmid. 2012. "Testing Dynamic Agency Theory via Structural Estimation." working paper, Duke Fuqua.

Rust, John. 2010. "Comments on Michael Keane's Structural versus Atheoretic Approaches to Econometrics." *Journal of Econometrics*.

Strebulaev, Ilya. 2007. "Do Tests of Capital Structure Mean what They Say?" *Journal of Finance* 62: 1747–1787.

Welch, Ivo. 2004. "Capital Structure and Stock Returns." *Journal of Political Economy* 112: 100–120.