Structural Models of Capital Structure:
A Framework for Model Evaluation and Testing

Arthur Korteweg
Michael Lemmon*

March 21, 2012

Abstract

Structural models of capital structure frequently have similar implications for the data that we observe, yet are typically not tested against alternative specifications. We examine the predictions of structural models, and develop a formal methodology for testing amongst alternatives. We show that our approach can provide a powerful yet simple decision-theoretic approach to model comparison, even for non-nested models. We show how our approach can highlight the dimensions on which various models fail, thus providing guidance to researchers for developing better models.

*Korteweg is at the Graduate School of Business, Stanford University, korteweg@stanford.edu. Lemmon is at the University of Utah, finmll@business.utah.edu.

Structural Models of Capital Structure:
A Framework for Model Evaluation and Testing

Abstract

Structural models of capital structure frequently have similar implications for the data that we observe, yet are typically not tested against alternative specifications. We examine the predictions of structural models, and develop a formal methodology for testing amongst alternatives. We show that our approach can provide a powerful yet simple decision-theoretic approach to model comparison, even for non-nested models. We show how our approach can highlight the dimensions on which various models fail, thus providing guidance to researchers for developing better models.

Introduction

The topic of capital structure remains an active area of financial research. While much has been learned about what factors are correlated with leverage and financing decisions, there still is no "workhorse" model of capital structure that is generally agreed upon. A more recent entrant has been the development of dynamic structural models of capital structure (e.g., Fisher, Heinkel, and Zechner (1989); Goldstein, Ju, and Leland (2001); Hennessy and Whited (2005), and Strebulaev (2007)). These models have been useful in highlighting how certain insights from static models do not necessarily generalize to a dynamic setting and by being able to offer potential explanations for some puzzling features of the data, such as the robust negative association between profitability and leverage and the weak evidence on rebalancing toward a target capital structure. Moreover, the structural approach, if the models are correct, also provides the ability to examine the effects of alternative policy choices, such as how changes in taxation or investment subsidies might affect the demand for various sources of capital. Nevertheless, while these models are appealing on a number of dimensions, the intuition driving their results remains opaque, and more importantly, in most cases, there has been no serious attempt to compare these models both to one another and to various alternative models that might be proposed. We seek to address some of these issues in this paper by first, examining the predictions of these models relative to some plausible alternative models along the existing dimensions upon which these models have been tested, and second, by developing a formal methodology for testing among alternative models.

On the first front, our results are not encouraging. Although these models can fit a number of moments of the data, we show that on these dimensions, they cannot generally be discriminated from simple alternative models, including a model of random financing choices. This is important, because it points out the difficulties associated with constructing powerful tests of alternative models and highlights some of the problems of the typical Simulated Method of Moments (SMM) approach. On the second, issue, we are more optimistic. We develop a framework for evaluating alternative structural models based on particle filtering and show that it can provide a powerful and simple decision-theoretic approach to model comparison, even for non-nested models. Moreover, we show how this approach can highlight the dimensions on which various models fail, thus providing guidance to researchers for developing better models. The remainder of the paper is organized as follows. In Section 2 we review a few of the existing models of capital structure. Section 3 summarizes the use of SMM in the existing studies. Section 4 describes our data. Section 5 presents a comparison of the existing structural models with some simple alternatives under the SMM approach. Section 6 introduces our framework for evaluation of alternative models, and Section 7 describes the empirical implementation using the particle filtering methodology. Section 8 presents empirical results using our approach. Section 9 discusses possible extensions and Section 10 concludes.

2. Structural Models of Capital Structure

All of the existing structural models of capital structure can be represented in the following basic state space form:

$$dx_{it} = \mu(x_{it}, t)dt + \sigma(x_{it}, t)dB_{it}, \qquad \text{(state evolution)}$$

$$y_{it} = f(x_{it};\, \phi) + \eta_{it}. \qquad \text{(observation equation)}$$

The state vector, x_{it}, is the sole source of the exogenous shocks in the model, and follows an Itô process with drift, $\mu(\cdot)$, and volatility, $\sigma(\cdot)$, both of which may depend on the current time, t, and the current value of the state for firm i at time t. At each point in time we observe a vector of outcome variables, y_{it}, which is a function of the underlying state and a set of exogenous parameters ϕ. The derivation of this function, $f(\cdot)$, is the bread and butter of the model. The observation noise, η_{it}, is not typically part of the model, but we include it for reasons that will become clear below. For notational convenience, we also define the parameter vector θ as the collection of the parameters in the state evolution (e.g. drift, volatility, speed of mean reversion parameters) and the observation equation, ϕ.

In Table 1, we summarize the main capital structure models in the literature by identifying the state variable, the observed variables, and the exogenous parameters in each model. For example, in the classic Merton (1974) model, the state variable is the value of the firm's assets, which follows a Geometric Brownian Motion, whereas in Goldstein, Ju and Leland (2001), x_{it} is the firm's EBIT. In other models the state is an unobserved variable. For example, in the Leland (1994) model the state is the value of the *unlevered* firm's assets. In Hennessy and Whited (2005, 2007) the state variable is the level of productivity. The vector y_{it} includes observed variables such as leverage, profitability, and capital expenditures, all of which are pinned down by the model based on the realization of the current state and the model's parameters.[1]

3. The SMM approach to model testing

The method of choice to date in evaluation of structural models of capital structure is Simulated Method of Moments (SMM). In the SMM approach, the researcher chooses a number of moments to match from the data. Then, starting with an initial guess for the values of the exogenous parameters, the researcher generates a set of simulated moments from the model under consideration and compares the values of the simulated moments to the moments in the actual data. A global optimization routine, such as simulated annealing, is then used to find the set of parameter values that minimizes the distance between the simulated and actual moments. A weighting matrix is generally used to weight the moments in a statistically optimal manner.

While the SMM approach is useful, its application in the literature suffers from several potential weaknesses. The first weakness is that models are not directly tested against each other. Rather

[1] Although the majority of the structural models are written in continuous time, some models (such as Hennessy and Whited (2005, 2007)) are written in discrete time. This does not change the subsequent discussion because the continuous-time models end up being discretized to fit data that is observed at regular time intervals. In other words, the empirical models that are taken to the data are all discrete-time versions of the theoretical models.

the J-test of over-identifying moment restrictions is used to judge whether a model is rejected by the data. Surviving the J-test does not mean that the model is “right” or that there is not another model that fits the data better. Conversely, what should we conclude if the J-test rejects the model? No benchmark model or alternative hypothesis is specified.

The second weakness of SMM is that it is not amenable to testing non-nested models. The literature typically reports significance tests for the individual model parameters. This is in essence a nested model test that informs us whether a restricted model fits the data equally well. However, structural models of the firm are typically non-nested, as is evident from the variety of state variables across models shown in Table 1, for example.[2] Moreover, even shutting down friction parameters in a given model does not always result in a meaningful benchmark. For example, turning off taxes and bankruptcy costs in Goldstein, Ju and Leland (2001) results in a model in which leverage drifts down to zero.

The third issue with SMM is that its application in the literature lacks statistical power. SMM does not require models to fit every observed data point, but only certain moments (e.g. means, variances, correlations) chosen by the researcher. It is common practice to estimate 8 or 9 parameters by matching 10 to 12 moments. Much of the dynamics in the data is lost and it is therefore more difficult to reject models. It is particularly important for capital structure models to use all available data in model testing, since different frictions can have very similar empirical predictions. For example, the well known negative relationship between leverage and profitability is consistent both with the financing hierarchy arising from asymmetric information as described by the pecking order (Myers and Majluf (1984)) and with dynamic tradeoff theories of capital structure that rely on the tradeoff between taxes and bankruptcy costs (Fisher, Heinkel, and Zechner (1989)).

[2] There is a small literature on testing non-linear, non-nested models using GMM. Based on the encompassing principle of Mizon and Richard (1986): Cox test, Davidson-McKinnon J-test. A notable exception is Singleton (1985) who considers a case where there is no encompassing model. However, the applications of these techniques are quite sparse due to the restrictive assumptions required. See also,Vuong (1989) and Hong and Preston (2011).

The fourth issue is that structural models of capital structure typically make predictions on y_{it}'s of different dimensions. For example, Leland (1994) only models the leverage decision whereas Hennessy and Whited (2005, 2007) also make predictions on investment. To our knowledge there is no established way to use SMM to compare models of different dimensionality.

The fifth and final limitation is that there is no theoretically founded adjustment for model complexity in SMM. Ideally we would like to find the "simplest" model that "fits" the data. Although the Wald test for joint parameter restrictions accounts for degrees of freedom, a more pertinent issue for model complexity is in the number of state variables and whether they are observed or latent. For example, with a latent state the researcher has much more flexibility to match the observed data than if the state is observed, and a model test should account for the higher degrees of freedom. With SMM this is not an easy task to accomplish.

A final issue, although not specific to SMM, is that the structural models are specified at the firm level and do not aggregate. Yet the extant literature aggregates firms to the level of the economy or certain broad subsamples of firms at best. For example, a well-known result in macro-economics is that although individual firms follow lumpy investment policies (e.g. due to fixed adjustment costs), aggregate investment may be smooth even when firms have identical parameters (Caballero and Engel, 1991). Hence, estimates based on aggregate moments may be misleading when the aim is to learn about firm-specific behavior. It is difficult to estimate firm-specific parameters with SMM because the available time series to calculate moments on a firm-by-firm basis is very short, making the issue of statistical power even more worrisome.

While SMM may be adapted to deal with some of the issues (such as increasing power by adding conditioning information and including more correlations to better capture dynamics), we propose a simple likelihood-based methodology that is solidly grounded in decision theory and deals with all of the problems that we identify above.

4. Data

To provide an illustration of our approach, we use a small set of firm specific data from Compustat. We first select firms that have 20 years of continuous data from 1989-2008, and find the median firm in the sample in terms of total book assets over the sample period. We then select the closest five firms (including the median firm). While this sample is clearly not representative, it provides us with a set of time-series at the firm level to evaluate our modeling procedure. One would expect that moving to the full sample of firms will only induce greater variation in the dynamics that we document here. We measure leverage following Hennessey and Whited (2005) as the total book value of long-term debt less the book value of cash all divided by the market value of the firm—defined as total assets less book equity plus market equity. To be consistent with the Goldstein, Ju, and Leland (2001) model (which we use as the benchmark structural model in this paper) we measure profitability (ROA) as earnings before interest, taxes, and depreciation scaled by the market value of the firm. We also compute the financing deficit as total net debt and equity issues net of payouts (interest and dividends, respectively).
Summary statistics of the two samples are reported in Table 2. Average book assets is $1.4 billion, with a standard deviation of $815 million. The large standard deviation is due to the change in assets over the 20 year sample period rather than a cross-sectional dispersion, as firms were selected to be of similar size. Average leverage is 0.076 and the standard deviation of leverage is 0.090. ROA is 0.112 on average, with a standard deviation of 0.037. Average Q is 2.044, and firms invest on average at a rate of 0.121, with a standard deviation of 0.101, showing that these firms grow considerably over the sample period. Average net equity issuance is -0.058, revealing that these firms are prodigious dividend payers. Average debt issuance is 0.006, with a standard deviation of 0.081. Figure 1, Panels A through F plot the key variables over time for a single firm (the median firm based on book assets). These figures illustrate the dynamics that we are interested in understanding.

5. SMM and model testing—An example

In this section we illustrate the difficulties in using SMM as a model evaluation tool. To do so, we fit four models to the representative data based on firm size above. The first is a random financing model in which the firm randomly decides to issue debt or equity to cover its external financing gap, with fixed probability p. The second is a simple pecking order inspired model in which the firm issues debt to cover its financing gap if leverage is below capacity (a parameter that is to be estimated), and equity if leverage is above capacity as in Lemmon and Zender (2010).[3] These two models are simple reduced-form models that we compare with the structural model. The third model and our first structural model is a variant of the Goldstein Ju and Leland (2001) tradeoff model as modified in Korteweg and Strebulaev (2010). The exogenous parameters in this model are: the risk free rate, the risk premium (used to move back and forth between the physical and risk-neutral measures), the refinancing cost (γ), the bankruptcy cost (α), and the volatility of the earnings process (σ). The fourth model is a Q theory model in the spirit of Hennessy and Whited (2005) with a linear tax structure. For all models we estimate the parameters using SMM. For the tradeoff model we attempt to match the mean and variance of both leverage and roa, while for the random financing model we fit only the mean and variance of leverage.

The results of the estimations are shown in Table 2 and are easy to summarize. At this point we make no attempt to assess the reasonableness of the parameter estimates. The tradeoff model fits leverage fairly well but has a particularly difficult time matching the volatility of the earnings process. This is because in this tradeoff model, the earnings process is a random walk with drift and does not allow for mean reversion. Thus firm value scales very closely with shocks to earnings making the ROA process less volatile. Most important is the comparison with the random financing model. As seen in the table, the moments for leverage generated via random financing are basically identical both the actual moments and to the moments from the tradeoff

[3] Note that this pecking order model is subtly different from a trade-off model in the sense that financing decisions are made incrementally in response to a need for external financing. In a trade-off model the firm should refinance if leverage drifts too far from optimum even if there is no need for external finance.

structural model. Simply put, at least based on matching the moments of leverage, the SMM tests have no power to reject one model over the other. While one could add additional moments that could potentially discriminate the two models (such as correlation between profitability and leverage), we take a different path based on a decision theoretic approach to testing non-nested models that exploits all of the information in the data. We introduce this approach in the following section.

6. A decision-theoretic approach to model testing

The goal of model testing is to use the observed data, Y, to compare two competing models, $\mathcal{M}_0$, and $\mathcal{M}_1$, that are possibly non-nested. The data set Y contains the observed variables y_{it} for all i and t, as well as the observed elements of the state, x_{it}, if any.

In a decision-theoretic framework, the researcher prefers model 1 if the expected utility of model 1 exceeds the expected utility from model 0:

$$\mathcal{U}_1 \cdot p(\mathcal{M}_1|Y) > \mathcal{U}_0 \cdot p(\mathcal{M}_0|Y)$$

where $\mathcal{U}_0$ is the utility from model 0 being true, and $p(\mathcal{M}_0|Y)$ the probability that $\mathcal{M}_0$ is the correct model based on the observed data.[4] Rewrite the decision rule to choosing model 1 if

$$\frac{p(\mathcal{M}_1|Y)}{p(\mathcal{M}_0|Y)} > \frac{\mathcal{U}_0}{\mathcal{U}_1}$$

In other words, the researcher chooses model 1 if the ratio of posterior model probabilities, called the *posterior odds ratio*, exceeds a threshold that is larger the higher the utility to model 0 relative to model 1. For example, the relative utilities may reflect the researcher's preference for a more complex model even if it fits the data slightly less well, because it provides more economic insight.

[4] It is tempting to equate $p(\mathcal{M}_0|Y)$ to the classical p-value, but this is not correct. The p-value is the probability of observing a dataset that is *at least* as "extreme" as the observed Y, under the null hypothesis. The p-value is therefore strongly biased against the null. In Jeffreys' (1939) famous words: "*What the use of P implies, therefore, is that a hypothesis that may be true may be rejected because it has not predicted observable results that have not occurred. This seems a remarkable procedure.*" We refer the interested reader to Berger and Delampady (1987) for a comprehensive treatment of p-values and hypothesis testing.

Applying Bayes' rule, express the posterior odds ratio as

$$\frac{p(\mathcal{M}_1|Y)}{p(\mathcal{M}_0|Y)} = \frac{p(Y|\mathcal{M}_1)}{p(Y|\mathcal{M}_0)} \cdot \frac{p(\mathcal{M}_1)}{p(\mathcal{M}_0)}$$

The first factor on the right-hand side is called the *Bayes factor* for $\mathcal{M}_1$ against $\mathcal{M}_0$, denoted by $\mathcal{B}_{10}$. A Bayes factor larger than one indicates that the data favors $\mathcal{M}_1$ over $\mathcal{M}_0$. The second factor is the prior odds ratio and is often set equal to one, indicating the absence of a prior preference for either model. The decision rule is then

$$\text{If } \mathcal{B}_{10} > c \Rightarrow \text{choose model 1.}$$

$$\text{If } \mathcal{B}_{10} \leq c \Rightarrow \text{choose model 0.}$$

The choice of threshold, $c \equiv \frac{\mathcal{U}_0 p(\mathcal{M}_0)}{\mathcal{U}_1 p(\mathcal{M}_1)}$, is rather subjective, depending on the prior odds ratio and the relative utility of the models. Jeffreys (1961) suggest applying a threshold of $c = 100$ as decisive evidence in favor of $\mathcal{M}_1$. We will report the Bayes factors and let the reader decide what constitutes strong evidence.

At first sight the Bayes factor looks like a classical likelihood ratio, but on closer inspection it becomes clear that it is a different object. While a likelihood is conditional on the model parameters, the marginal likelihoods, $p(Y|\mathcal{M}_j)$, that make up the Bayes factor integrate out the parameters from the likelihood:[5]

$$p(Y|\mathcal{M}_j) = \int p(Y|\theta, \mathcal{M}_j) p\left(\theta|\mathcal{M}_j\right) d\theta$$

The advantage of integrating out the parameters is that the Bayes factor functions like a fully automated Occam's razor: it trades off model fit and parsimony in a theoretically consistent way. Figure 2 illustrates the intuition. Consider two restricted, parsimonious models that fix the parameter θ to either θ_1 or θ_2. Their likelihoods are plotted in Figure 2 with a striped and dotted

[5] The following is a slight abuse of notation. Since the parameters technically depend on the model specification, the parameter should be labeled as θ^$\{\mathcal{M}_j\}$. To avoid overly complex notation, we drop the dependence on the model.

curve, respectively. Now consider the unrestricted, more complex model, where θ can potentially take on either value (and hence θ needs to be estimated for this model). Its (integrated) marginal likelihood, $p(Y)$, is a weighted average of the likelihoods of the restricted models, and plotted as a solid line in Figure 2. Relative to the restricted models, the complex model is more flexible: its likelihood is “stretched” over the support and can therefore fit many realized datasets with reasonable probability. However, for any given dataset, one of the restricted models always fits the data better. In other words, the flexibility of the complex model is a double-edged sword, as it can fit many datasets a priori, but fits no dataset particularly well and is therefore punished for its flexibility when comparing it to more parsimonious models, which have sharper predictions about what datasets are expected to be observed.

This intuition extends to non-nested models and models with latent state variables, where in the latter case the integration is over both the model parameters and the latent states. Interpreting a latent state as a (somewhat restricted) sequence of free parameters, the state variable affords great flexibility in fitting the observed data but also makes the model more difficult to validate relative to models with observed states.

Thus far it was implicitly assumed that all models have predictions on the same variables, such that Y is the same across models. There are two ways to deal with models that make predictions on different sets of variables. The first is to integrate out the dimensions of Y that are not common to the models considered. This allows models to be compared on an equal basis, but has the disadvantage of ignoring the richness of predictions that more sophisticated models have. The second way is to specify priors on the dimensions that are not modeled (e.g. investment in Leland’s (1994) model). The latter approach carries the same intuition as the parsimony-fit trade-off: models that do not make predictions on certain dimensions are penalized (assuming diffuse priors) because those dimensions are integrated out in order to compute $p\left(Y|\mathcal{M}_j\right)$.

The Bayes factor is conditioned on all observed data and is generally a statistically more powerful approach than SMM: it is easier to reject a given model when every data point needs to be explained rather than matching a select set of moments derived from the data. Moreover, each firm may be fitted individually so there is no need to average across companies to fit aggregate moments. However, the approach does come at the expense of making stronger distributional assumptions about the observation errors, an issue to which we will return to below.

The above insights generalize to model testing in the presence of more than two models. For example, we can pick a baseline model and compare all models against the baseline, or alternatively, compute relative model probabilities within the set of models considered. The set of models need not be exhaustive (i.e. $\sum_j \; p\left(\mathcal{M}_j | Y\right)$ need not equal one), as we are purely comparing models against each other.

7. The Particle MLE implementation of model testing

We use the Bayesian Information Criterion (BIC, also known as the Schwarz criterion) to approximate Bayes factors, using a Maximum Likelihood Estimator (MLE). This approach is effective in dealing with the dimensionality of structural models while preserving the advantages of the Bayes factor outlined above.

7.1. The Bayesian Information Criterion

A useful asymptotic approximation to the marginal probability $p\left(Y | \mathcal{M}_j\right)$ can be derived using Laplace's method for integrals (see appendix A for a detailed derivation):

$$p\left(Y | \mathcal{M}_j\right) \approx p\left(Y | \hat{\theta}, \mathcal{M}_j\right) - \frac{d_j}{2} \log s$$

Where $\hat{\theta}$ is the MLE of the model parameters, d_j is the number of parameters in model j, and s is the total number of firm-years. This particular approximation assumes Normally distributed priors on the parameters, with the same information content as a single observation.

The above expression shows that we can use the likelihood of the fitted model, $p\left(Y|\hat{\theta}, \mathcal{M}_j\right)$, to approximate the Bayes factor. Define

$$BIC_j \equiv -2\log p\left(Y|\hat{\theta}, \mathcal{M}_j\right) + d_j \log s$$

The relation between BIC and the Bayes factor is then

$$2\log \mathcal{B}_{10} \approx BIC_0 - BIC_1$$

The approximation error is of order $O\left(s^{-\frac{1}{2}}\right)$, and thus tends to zero as s grows larger.

7.2. Estimating the likelihood

If the state variables are all observed then maximizing the likelihood (and hence computing the BIC) is a standard MLE exercise. With an unobserved state vector, however, the likelihood function requires us to integrate out the latent state variables:

$$\log p(Y|\theta) = \log p(y_1 \dots y_T|\theta)$$

$$= \sum_{t=1}^{T} \log \mathrm{p}(\mathrm{y_t}|\mathrm{y^{t-1}}, \theta)$$

$$= \sum_{t=1}^{T} \log \int p(y_t|x_t, \theta) p(x_t|y^{t-1}, \theta) dx_t$$

where $y^T = \{y_1 \dots y_T\}$. We suppress the conditioning on the model, $\mathcal{M}_j$, for ease of exposition. Given the non-linearity of the observations in the states, we use the particle filter to evaluate the integral. The particle filter is fast and does not suffer from the dimensionality problems of importance sampling in Simulated Maximum Likelihood, where the variance of the importance weights rises rapidly as the length of the time series increases (Pitt, 2002).

The particle filter is a non-linear filtering method used to obtain a discretized approximation to the posterior distribution of the state, $p(x_t|y^t, \theta)$, and to update it sequentially as new data arrives. The filter relies on two basic requirements, namely the ability to: 1) simulate forward the state vector, and; 2) evaluate the conditional likelihood $p(y_t|x_t, \theta)$. Note that these two steps are also present in the integral above, underscoring that this a natural approach to evaluating the integral. From the assumption that the state follows an Itô process, step (1) is straightforward. With additive observation errors, η, step (2) also does not pose any problem.

Figure 3 illustrates the basic the intuition behind the particle filter. The top left corner shows the discrete approximation to the distribution of the state at time t using a number of particles (points in the state space), with associated probability weights. The first step simulates each particle forward by one period using the state evolution process. This simulation yields a sample of the predictive density $p(x_{t+1}|y^t, \theta)$, but does not account for the new information in y_{t+1}. The second step therefore reweighs the particles proportional to the conditional likelihood $p(y_{t+1}|x_{t+1}, \theta)$. The resulting distribution approximates the updated filtering density $p(x_{t+1}|y^{t+1}, \theta)$.

Appendix B describes the particle filter in more detail and explains how to use the particle approximation to evaluate the likelihood.

8. Empirical results using particle MLE

In this section, we estimate each of the models described above using particle MLE. We use the data on five representative firms based on firm size (total book assets) and fit the models to the 100 data points.

8.1. Summary Statistics and Correlations

To provide some initial evidence on how well the models fit the data Table 4 compares a number of moments from the data with the predicted moments from the various models. To be specific, the predictive moments are computed based on the one-step ahead forecasts. The first row of the table reports the results for leverage. As seen in the table, all four of the models provide reasonable fits to both the mean and standard deviation of leverage. Note, however, that the simpler random financing and pecking order models actually match the data more closely. This is not too surprising, given that these models have little to say about other moments, such as investment and profitability. Nevertheless, it is a useful reminder that even simple models can easily match some of the primary features of the data. Similarly, for debt and equity issuances none of the models exhibit an obvious superiority. Finally, comparing the correlations between leverage and debt and equity issuance from the data with those from the models, the pecking order model most closely matches the data, while all of the other models get at least one of the signs of the correlations incorrect.

Next comparing the dynamic tradeoff model and the Q model with the data on profitability, it can be seen that the Q model does a better job of capturing the variability of profitability. The fact that the tradeoff model generates little variation in profitability is an artifact of the fact that in this model assets and profitability both grow in proportion to the shocks resulting in constant ROA across time. Both models generate a strong negative relationship between profitability and leverage, but in both cases, the correlation is much stronger than that exhibited in the data (which is actually positive in this particular sample).

Finally, the Q model predicts investment that is higher and more variable than that observed in the data and also produces correlation between investment and leverage, debt issuance and equity issuance that are higher than those observed in the data.

Next, in Table 5, we report the correlations between the predictive means from the models and the realized data. These correlations provide some initial information on how well the models do at capturing the dynamics of the data. For leverage, the random financing model produces a correlation with the data of 0.61, which is similar to that of the pecking order model (0.57). In contrast the correlations of predicted leverage with the data for the tradeoff and Q models are 0.34 and 0.30, respectively. In terms of profitability, both the tradeoff model and the Q model exhibit low correlations with the data. Similarly, predicted investment from the Q model is negatively correlated with actual investment. Finally, all models exhibit reasonably large correlations between predicted and actual debt issuance, but only the random financing and pecking order models have correlations that are the right sign for predicted and actual debt issuance.

Table 6 displays the average sources and uses of funds from the data and compares these with the predicted moments from the Q model. On average, the Q model overpredicts the fraction of sources of funds coming from debt issuance, but generates higher levels of capex and interest expense/debt repayment compared to the data.

8.2 Likelihood Analysis

Figure 4 displays the likelihood analysis based on the particle filter estimation for the four models based on leverage only. As seen in the top and middle panels of the table, both the random financing and pecking order models have higher likelihood values compared to the more complex tradeoff and Q models in every single period. This suggests that the better fit of the simpler models is not due to any particular time period but rather is uniform over the sample. In contrast, the better fit of the tradeoff model is driven mostly by the pre-2003 period (boom vs bust?) The last panel in the Figure, shows that, based on leverage alone, according to the BIC, one would favor the simpler models (random financing and pecking order) versus the more complex models. This is consistent with the data in Section 8.1, which showed that the simple models can easily match average leverage as well as the correlations between leverage and issuance. The next figure compares the models along the dimensions of both leverage and ROA jointly. In order to make this comparison, we augment the random financing and pecking order models with a model for ROA in which ROA is assumed to be drawn from an i.i.d. Normal distribution with the empirical mean and variance.[6] This is a fair comparison to the tradeoff model, because the tradeoff model predicts that ROA is a constant when the frictions (taxes and bankruptcy costs) are turned off. The i.i.d. variance can then be interpreted as the noise component. We get very similar results if we use an AR(1) model for ROA instead, which is close the process implied by the Q

[6] Note that the distribution of ROA must be proper for BIC to work (Raftery (1995)).

model (we do not report the these results). As seen in this figure, the more complex models now perform better, but the simple models are still at least as good according to BIC.

In Table 7, we compute the various contributions to the log-likelihood of the various models. The table reports the base case (where the predictive moments are used) and the "in Sample" case where it is assumed that the state is known. Comparing the base case with the in sample likelihoods, the main take away is that the likelihoods of models are dramatically improved if one knows the state in the current period rather than having to forecast it out-of-sample.

Finally, we provide some assessment of how well the various models do in capturing the dynamics in the data. For example, does the firm in the Q model issue debt or increase leverage at the correct times? Figure 6 displays the data for the dynamics of leverage for each of the five firms used in the estimation. Examining the figures, the following features are evident. First, the simple models do a good job of tracking actual leverage, whereas both the tradeoff model and to an even greater extent, the Q model exhibit much smoother dynamics than the actual data. This is intriguing, because it suggests that the frictions in these models appear to cause firms to smooth leverage to a much greater extent than what appears in the data. The second figure repeats the analysis above but based on the in sample fitted values. Overall the results are similar with the exception of the fact that the dynamic tradeoff model now tracks the evolution of leverage more closely, while the Q model still exhibits muted dynamics compared to the actual data.

The next figure compares the dynamics of ROA from the data with ROA implied by the tradeoff and Q models. Again, the main feature evident in the figures is that the models have a difficult time capturing the actual dynamics in the data, suggesting that some source of additional variation in profitability is needed to improve the model fit.[7]

[7] Note that this does not mean that the tax-bankruptcy cost model is invalidated, it just means that this particular model does not fit the data well. This is informative, though, because it tells us that an improvement to the GJL model needs to decouple EBIT and equity market value in order to fit the data better.

Finally, Figure 9 shows the dynamics of investment from the Q model compared to the data. The figures are interesting. While the model delivers investment with similar or even greater variability than in the actual data, the dynamics are not similar. This is true both when one uses the predictive moments or the in sample moments and indicates that investment in the data is driven by something beyond simple productivity shocks. Better understanding these dynamics represents a challenge for future research.

9. Discussion and possible extensions

In this section, we discuss a number of implications of our analysis for future research on structural models and their application to capital structure.

9.1 Interpretation of results and guidelines for future structural work

First, the bulk of structural models in the literature only incorporate cash flow shocks but do not allow for discount rate shocks. This has large implications for empirical fit, as seen particularly clearly in the tradeoff model, where ROA does not exhibit any time variation. Recent models such as Chen (2010) and Bhamra et al. (2010) have started to incorporate pricing kernels into these models, and our results suggest that this is a fruitful direction.

Second, the Q model lacks "real" frictions such as adjustment costs. The result is that investment (and hence financing) closely follows Q (and hence shocks to market values), resulting in too much issuance activity relative to the observed data. For non dividend-payers financial frictions (issuance costs) may take on the role of real adjustment costs. This does not happen in our sample of firms because they tend to pay dividends, and the distribution tax lowers the threshold to invest.

Third, models need to be compared on dimensions other than attributes like the mean and variance of leverage, and more formal comparisons against alternative models should be made. Our results show that naïve models can fit a number of moments in the data and even yield dynamics that are similar to those observed.

Finally, structural models are a combination of economic and statistical assumptions. The economics should be center-stage, but the statistical assumptions can have important empirical implications. For example, Kane, Marcus and McDonald (1983) showed that one needs a jump in the driving process in order to explain the simultaneous existence of levered and unlevered firms.

9.2 Methodology

The basic trade-off between SMM and Likelihood-based methods is between having a more powerful estimator and making stronger distributional assumptions on the observation errors, η. However, the difference in distributional assumptions between our approach and SMM is not as large as it may seem at first sight. In SMM you still need the distributional assumption on the state diffusion. The difference is that in likelihood-based methods you also need to make assumptions about the observation error distribution. The distributional assumptions on observation error can be relaxed through using mixtures of Normals.

Another perspective on the relation between the two methods is that MLE is essentially doing SMM on the score, and we treat each firm-year as its own separate moment. In other words, we push SMM to consider many more moments.

Note: you need the observation errors (η's), without them you run into a stochastic singularity: since no model can perfectly fit every datapoint, without the observation errors we can simply reject every single model. We therefore need to allow for some model misspecification. The variance of η indicates how closely the model fits the data.

Heterogeneity in parameters can be accommodated through hierarchical priors (akin to random effects).

We ignore predictions of theories outside of the model. We also ignore characteristics of the data outside of the models (e.g. the robust relation between size and leverage in the data is not captured by any of the models we considered).

More deeply, we should warn against using our approach as the sole guiding principle in comparing models. The goal should not be to set out on a fishing expedition to find the model that best fits the data.

- Cannot distinguish how “close” the model’s results are to its assumptions.
- Cannot distinguish economic vs statistical/distributional assumptions.
- Can always cook up a reduced-form model that outperforms a structural model? Related point: can always find a simple model that works better in-sample (see Fig 1)
    - Importance of out-of-sample testing!!
- Other definitions of parsimony/complexity
    - e.g. existence of closed-form solution (whereas our measure reflects complexity in a statistical sense, i.e. # parameters and state variables).

Rather, our methodology is useful to diagnose where models do and do not fit the data and to serve as input into a broader discussion of the relative merits of a model.

10. Conclusion

We present a general approach for comparing the empirical fit of state space models. Our approach is motivated from a Bayesian perspective on relative model probabilities, and relies on maximum likelihood estimation with particle filtering that is effective in dealing with the complex nature of structural models in corporate finance. These models are typically non-linear, non-nested and are driven by unobserved state variables, making them difficult not only to estimate but also to compare across models. In addition, we present two strategies for making comparisons across models that make predictions on different sets of variables that may only partially overlap.

Our framework allows us to formally contrast the empirical fit of models in both a predictive and an in-sample sense, and to consider the full dynamics of the data rather than a selected set of moments.

We show that with standard SMM/GMM methods it is difficult to distinguish models based on their fit to leverage alone. Using our framework instead, we find that simple random financing and pecking order models fit the leverage data better than a canonical tax-bankruptcy cost trade-off structural model or a Q-theory based investment model with financing frictions. However, this should not be taken as a defeat of the structural models, as they make rich predictions on variables other than leverage, unlike the simple models. For example, when we augment the simple models with a model for profitability, the difference between the simple and structural models shrinks considerably.

Based on our empirical findings we provide a number of suggestions for improving the structural models going forward: First incorporating discount rate shocks in addition to cash flow shocks will help to generate market value movements that are closer to those observed in the data while keeping investment relative stable. Second, introducing "real" frictions such as adjustment costs to changing the productive capital stock will result in more realistic financing patterns. Finally, the specification of the driving process in the structural models has a first-order effect on the empirical fit. It is important to recognize and diagnose the effect of the statistical assumptions relative to the economic assumptions that make up a structural model.

Finally, due to the state-space representation, our methodology can also be applied to asset pricing and DSGE models in macro-economics.

References

Berger, James O. and Mohan Delampady, 1987, Testing precise hypotheses, *Statistical Science* 2, 317-335.

Caballero, Ricardo J. and Eduardo M.R.A. Engel, 1991, Dynamic (S,s) economies, *Econometrica* 59, 1659-1686.

Chang, Xin and Sudipto Dasgupta, 2009, Target behavior and financing: How conclusive is the evidence? *Journal of Finance* 64, 1767-1796.

Fischer, Edwin O., Robert Heinkel, and Josef Zechner, 1989, Dynamic capital structure choice: Theory and tests, *Journal of Finance* 44, 19–40.

Gordon, Neil.J., D.J. Salmond, and Adrian .F.M. Smith, 1993, A novel approach to non-linear and non-Gaussian Bayesian state estimation, *IEE Proceedings F* 140, 107-133.

Jeffreys, Harold, 1961, *Theory of Probability* (Oxford University Press, Oxford).

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.

Leland,Hayne, 1994, Risky debt, bond covenants and optimal capital structure, *Journal of Finance* 49, 1213–1252.

Lemmon, Michael L. and Jaime F. Zender, 2010, Debt capacity and tests of capital structure theories, *Journal of Financial and Quantitative Analysis* 45, 1161-1187.

Mauer, David C. and Alexander J. Triantis, 1994, Interactions of corporate financing and investment decisions: A dynamic framework, *Journal of Finance* 49, 1253-1277.

Merton, Robert C., 1974, On the pricing of corporate debt: The risk structure of interest rates, *Journal of Finance* 29, 449–470.

Pitt, Michael K., 2002, Smooth particle filters for likelihood evaluation and maximisation, *working paper*, University of Warwick.

Pitt, Michael K. and Neil Shephard, 1999, Filtering via simulation based on auxiliary particle filters, *Journal of the American Statistical Association* 94, 590-599.

Raftery, Adrian.E., 1995, Bayesian model selection in social research, *Sociological Methodology*, 25, 111-196.

Rubin, Donald B., 1988, Using the SIR algorithm to simulate posterior distributions. In: Bernardo, J.M., DeGroot, M.H., Lindley, D.V., & Smith, A.F.M., eds., *Bayesian Statistics 3*, 395-402, Oxford University Press, Oxford.

Strebulaev, Ilya A., 2007, Do tests of capital structure theory mean what they say? *Journal of Finance* 62, 1747–1787.

Luke Tierney and Joseph Kadane, 1986, Accurate approximations for posterior moments and marginal densities, *Journal of the American Statistical Association* 81, 82-86.

Appendix A - Derivation of the BIC approximation to the Bayes factor

This derivation follows Raftery (1995), using Laplace's method to approximate the integral in the marginal likelihood,

$$p(Y|\mathcal{M}_j) = \int p(Y|\theta, \mathcal{M}_j) p\left(\theta|\mathcal{M}_j\right) d\theta$$

Define $g(\theta) \equiv \log\left[p(Y|\theta, \mathcal{M}_j) p(\theta|\mathcal{M}_j)\right]$ and consider the Taylor series expansion around its maximum, $\tilde{\theta}$:

$$g(\theta) = g(\tilde{\theta}) + (\theta - \tilde{\theta})' g'(\tilde{\theta}) + \frac{1}{2}(\theta - \tilde{\theta})' g''(\tilde{\theta})(\theta - \tilde{\theta}) + o\left(\|\theta - \tilde{\theta}\|^2\right)$$

By definition of $\tilde{\theta}$, its vector of first derivatives, $g'(\tilde{\theta})$, equals zero, so

$$g(\theta) \approx g(\tilde{\theta}) + \frac{1}{2}(\theta - \tilde{\theta})' g''(\tilde{\theta})(\theta - \tilde{\theta})$$

The Taylor expansion works well only for values of θ close to $\tilde{\theta}$. However, with a large number of observations, the likelihood is concentrated around the maximum and declines fast as one moves away from $\tilde{\theta}$. Hence, only values close to $\tilde{\theta}$ will contribute to the integral (Tierney and Kadane (1986) provide a formal argument). Plug the Taylor expansion into the integral:

$$p(Y|\mathcal{M}_j) = \int \exp(g(\theta))\, d\theta$$

$$\approx \exp\left(g(\tilde{\theta})\right) \int \exp\left(\frac{1}{2}(\theta - \tilde{\theta})' g''(\tilde{\theta})(\theta - \tilde{\theta})\right) d\theta$$

Since the integrand is proportional to the multivariate Normal density, it integrates to a constant,

$$p(Y|\mathcal{M}_j) \approx \exp\left(g(\tilde{\theta})\right) (2\pi)^{d_j/2} |A|^{-1/2}$$

where d_j is the number of parameters of model j (i.e. the dimension of the vector θ) and $A \equiv -g''(\tilde{\theta})$, the negative of the Hessian matrix of second derivatives. The error in this equation is of order $O(s^{-1})$, with s the number of observations. Hence,

$$\log p(Y|\mathcal{M}_j) = \log p(Y|\tilde{\theta}, \mathcal{M}_j) + \log p\left(\tilde{\theta}|\mathcal{M}_j\right) + \left(\frac{d_j}{2}\right) \log 2\pi - \frac{1}{2}\log|A| + O(s^{-1})$$

In large samples $\tilde{\theta}$ converges to the MLE, $\hat{\theta}$, and $A \approx s \cdot i$, where i is the expected Fisher information matrix for a single observation, a $d_j \times d_j$ Hessian matrix

$$i = -E\left[\left.\frac{\square^2 \log p\left(y_t|\theta, \mathcal{M}_j\right)}{\square\theta\square\theta'}\right|_{\theta=\hat{\theta}}\right]$$

with the expectation taken over y_t. Therefore, $|A| \approx s^{d_j} \cdot |i|$. Since these approximations are of order $O\left(s^{-1/2}\right)$,

$$\log p\left(Y|\mathcal{M}_j\right)$$

$$= \log p\left(Y|\hat{\theta}, \mathcal{M}_j\right) + \log p\left(\hat{\theta}|\mathcal{M}_j\right) + \left(\frac{d_j}{2}\right)\log 2\pi - \frac{d_j}{2}\log s - \frac{1}{2}\log|i| + O\left(s^{-1/2}\right)$$

Assume that the prior distribution $p\left(\theta|\mathcal{M}_j\right)$ is multivariate Normal with mean $\hat{\theta}$ and variance matrix i^{-1}, i.e. the prior distribution contains roughly the same information as a single observation, on average. Thus

$$\log p\left(\hat{\theta}|\mathcal{M}_j\right) = -\frac{d_j}{2}\log 2\pi + \frac{1}{2}\log|i|$$

and

$$\log p\left(Y|\mathcal{M}_j\right) = \log p\left(Y|\hat{\theta}, \mathcal{M}_j\right) - \frac{d_j}{2}\log s + O\left(s^{-1/2}\right)$$

The error term thus tends to zero as the sample size becomes larger.[8]

Define the Bayes Information Criterion (BIC) for model j as

$$BIC_j \equiv -2\log p\left(Y|\hat{\theta}, \mathcal{M}_j\right) + d_j \log s$$

so that

$$BIC_j = -2 \log p\left(Y|\mathcal{M}_j\right) + O\left(s^{-1/2}\right)$$

The BIC can be used to approximate the Bayes factor

$$2\log \mathcal{B}_{10} = 2\log p\left(Y|\hat{\theta}, \mathcal{M}_1\right) - 2\log p\left(Y|\hat{\theta}, \mathcal{M}_0\right) - (d_1 - d_0)\log s + O\left(s^{-\frac{1}{2}}\right)$$

[8] For other prior distributions, the error term is and does not vanish asymptotically, but the other terms eventually dominate as s grows large. Thus, the error tends to zero as a proportion of .

$\approx BIC_0 - BIC_1$

Appendix B - Particle filter and Likelihood evaluation.

Consider the discretized approximation to the filtering distribution $p(x_t|y^t)$ using a large sample of "particles", $x_t^1 \ldots x_t^M$, with probability weights $\pi_t^1 \ldots \pi_t^M$. Since we only use filters for which the weights are flat, we simply substitute in $1/M$ for the weights. We also suppress the conditioning on the model, $\mathcal{M}$, and its parameters, θ, for ease of exposition. The particle approximation to the filtering density is:

$$p(x_t = x|y^t) \approx \frac{1}{M} \sum_{k=1}^{M} \mathbb{I}\{x_t^k = x\}$$

where $\mathbb{I}\{\cdot\}$ is the indicator function. In words, the probability of $x_t = x$ is simply the proportion of particles for which $x_t^k = x$.

This appendix describes two popular methods to sequentially update the filtering distribution, and explains how to use the particle approximation to evaluate the likelihood.

SIR particle filter

The Sampling/Importance Resampling (SIR) filter of Rubin (1988) and Gordon et al. (1993) follows three steps to update the filtering density:

Algorithm: SIR Particle Filter

1) Randomly draw particles (with replacement).
2) Simulate each particle forward using the state evolution: , for .
3) Compute weights , for .
4) Resample M particles using the weights, normalized to sum to one, as the probabilities.

After resampling, the M particles again have weights $1/M$, and form an approximation to the density $p(x_{t+1}|y^{t+1})$. This approximation then feeds into step 1 for next period's update. With $R \to \infty$ the approximation $\frac{1}{R}\sum_{j=1}^{R} w^j$ converges in probability to the conditional likelihood, $\mathrm{p}(\mathrm{y}_{\mathrm{t}+1}|\mathrm{y}^{\mathrm{t}}, \theta)$.

Auxiliary particle filter

The SIR algorithm, although very general, is not the most efficient because it makes blind proposals of x_{t+1} in the first step, ignoring the fact that y_{t+1} is known. This is particularly problematic if the likelihood is highly peaked (or, more generally, when the w^j are highly variable). Intuitively, many particles will end up having zero weight and the resampling is based on only a few particles. An *adapted* filter such as the auxiliary particle filter of Pitt and Shephard (1999) makes more efficient use of the particles. A popular implementation of this algorithm uses the conditional mean of the state evolution, $\mu_{t+1}^j = \mathbb{E}\left[x_{t+1}|x_t^j\right]$ for each particle j:

Algorithm: Auxiliary Particle Filter

1) Compute first-stage weights for .
2) Draw R particle indices for , from a Multinomial distribution with probabilities , normalized to sum to one.
3) Simulate R realizations of from the state evolution: , for .
4) Compute second-stage weights , for .
5) Resample M particles using the second-stage weights, normalized to sum to one, as the probabilities.

The efficiency gain comes from the first stage, which ensures that we draw particles that have large predictive likelihoods. Setting all $g^m = 1$ collapses the algorithm back to the SIR filter.

Evaluating the likelihood

Pitt (2002) shows that an efficient and unbiased estimate of the conditional likelihood is the product of the average first and the second-stage weights (both unnormalized):

$$\hat{\mathrm{p}}(\mathrm{y}_{t+1}|\mathrm{y}^t, \theta) = \frac{1}{M}\sum_{m=1}^{M} g^m \cdot \frac{1}{R}\sum_{j=1}^{R} w^j$$

For the SIR the first stage weights are equal to unity, so $\frac{1}{M}\sum_{m=1}^{M} g^m = 1$, and the likelihood estimate is simply the average of the weights w^j, as argued above.

Table 1: Structural models of capital structure

Summary of a subset of the capital structure models in the literature. The canonical state space model for firm i at time t is

$$dx_{it} = \mu(x_{it}, t)dt + \sigma(x_{it}, t)dB_{it}, \qquad \text{(state evolution)}$$

$$y_{it} = f(x_{it};\ \phi) + \eta_{it}. \qquad \text{(observation equation)}$$

The table identifies the state variable (x_{it}), the form of the state evolution, the observed variables (y_{it}), and the exogenous parameters (ϕ) in the observation equation.

Model	x_{it}	State evolution	y_{it}	ϕ
Merton (1974)	Assets market value	Geometric Brownian Motion	Debt market value Equity market value	Risk-free rate Debt face value Debt maturity
Mauer and Triantis (1994)	Commodity price	Geometric Brownian Motion	Debt coupon rate Net debt issuance Production decision (yes or no)	Risk-free rate Corporate tax rate Bankruptcy cost Recap cost (fixed and proportional) Production costs Investment horizon Salvage value Convenience yield Operating adj. cost
Leland (1994)	Unlevered assets market value	Geometric Brownian Motion	Debt coupon rate Debt market value Equity market value	Risk-free rate Corporate tax rate Bankruptcy cost
Goldstein, Ju and Leland (2001)	EBIT	Geometric Brownian Motion	Debt coupon rate Debt market value Equity market value	Risk-free rate Corporate tax rate Personal tax rate Bankruptcy cost
Hennessy and Whited (2005)	Productivity shock	Discrete-time Gaussian AR(1)	EBIT Debt face value Assets book value Equity market value Net equity issuance Capital expenditures	Risk-free rate Corporate tax rate schedule (2 param's) Dividend tax rate Interest tax rate Depreciation rate Fire sale cost EBIT return to scale

Table 2: Summary statistics

Summary statistics for data used in the paper. We selected five firms around the median book assets in the data between 1989 and 2008. Book asset is in millions of dollars. Leverage is the book value of long-term debt less minus cash scaled by the market value of the firm (defined as total assets less book equity plus market equity). ROA is earnings before interest, tax and depreciation scaled by the market value of the firm. Q is the market value of equity plus book debt divided by book equity plus book debt. Investment is capital expenditures divided by property, plant and equipment. Equity issuance is defined as net equity issuance (issuance minus repurchases) less dividends, scaled by property, plant and equipment. Debt issuance is net debt issuance (issuance minus repurchases) less the change in cash balances, divided by property, plant and equipment. Source: Compustat.

	Mean	Median	St. Dev	Min	Max
Book assets	1338.6	1174.2	815.4	324.0	4746.4
Leverage	0.076	0.080	0.090	-0.129	0.306
ROA	0.112	0.105	0.037	0.041	0.239
Q	2.044	1.774	0.951	0.933	5.382
Investment	0.121	0.101	0.077	0.020	0.388
Equity issuance	-0.058	-0.054	0.049	-0.253	0.025
Debt issuance	0.006	0.002	0.081	-0.249	0.250

Table 3: SMM fit

Comparison of fit to moments for Random Financing Model and Goldstein, Ju, and Leland Model.

Actual Moments			
Mean Leverage	Stdev Leverage	Mean ROA	Stdev ROA
0.207	0.098	0.096	0.034
Estimated Parameters (Tradeoff Model)			
γ (refinancing cost)	α (bankruptcy cost)	Risk Premium	σ (earnings volatiltity)
0.047	0.631	0.033	0.230
Simulated Moments (Tradeoff Model)			
Mean Leverage	Stdev Leverage	Mean ROA	Stdev ROA
0.2105	0.127	0.0761	0.002
Estimated Parameters (Random Financing Model)			
P (prob of debt issue)			
0.888			
Simulated Moments (Random Financing Model)			
Mean Leverage	Stdev Leverage	Mean ROA	Stdev ROA
0.203	0.099	NA	NA

Table 4: Predictive Moments

'Data' column reports the realized moments in the data. The standard deviations are averages of the within-firm standard deviations.

	Data	Random Financing	Pecking Order	Trade-off	Q model
Panel A: First and second moments					
Mean(Lev)	0.076	0.074	0.074	0.080	0.089
Std(Lev)	0.053	0.055	0.056	0.033	0.031
Mean(ROA)	0.112			0.099	0.123
Std(ROA)	0.030			0.000	0.013
Mean(Q)	2.044				3.056
Std(Q)	0.566				0.655
Mean(Inv)	0.121				0.263
Std(Inv)	0.063				0.236
Mean(Eq iss)	-0.058	0.004	0.002	0.000	-0.159
Std(Eq iss)	0.035	0.052	0.071	0.000	0.081
Mean(Debt iss)	0.006	0.003	0.004	0.010	0.116
Std(Debt iss)	0.080	0.044	0.040	0.006	0.200
Panel B: Correlations					
Corr(ROA,Lev)	0.167			-0.997	-0.536
Corr(Inv,Q)	0.174				0.910
Corr(Inv,Lev)	0.159				0.576
Corr(Inv,Eqiss)	-0.138				0.210
Corr(Inv,Debtiss)	0.388				0.921
Corr(Lev,Eqiss)	-0.058	0.202	-0.047	-0.181	0.462
Corr(Lev,Debtiss)	0.360	0.202	0.239	-0.458	0.414

Table 5: Correlations between Predictive Means and Realized Data

Correlations between predictive means and realized data

	Random Financing	Pecking Order	Trade-off	Q Model
Lev	0.610	0.573	0.339	0.305
ROA			-0.158	0.092
Q				-0.130
Investment				-0.112
Equity issuance	0.294	0.327	-0.017	-0.073
Debt issuance	0.360	0.191	0.226	0.246

Table 6: Sources and Uses of Funds

All numbers are averages across firms-years, and scaled by beginning-of-period PPE.

	Data	Q model
Sources:		
Profits (ebit - tax):	0.18	0.29
Equity issuance:	0	0
Debt issuance:	0.03	0.16
Total:	0.21	0.46
Uses:		
CapEx:	0.12	0.26
Dividends+Share repo:	0.03	0.05
Interest+Debt repayment:	0.06	0.16
Total	0.21	0.47

Table 7: Breakdown of Log-likelihood

Breakdown of log-likelihood.

	Base case				"In-sample"	
	Random Financing	Pecking Order	Trade-off	Q model	Trade-off	Q model
Lev	162.41	159.72	50.69	-26.61	209.24	147.74
ROA(mkt)			94.87	79.83	264.88	237.89
Q				-368.52		8.06
Inv				-126.66		92.75
Net equity issue				-1,950.81		123.78
Total	162.41	159.72	320.13	-10,194.7	3,682.09	709,652.9

Figure 1

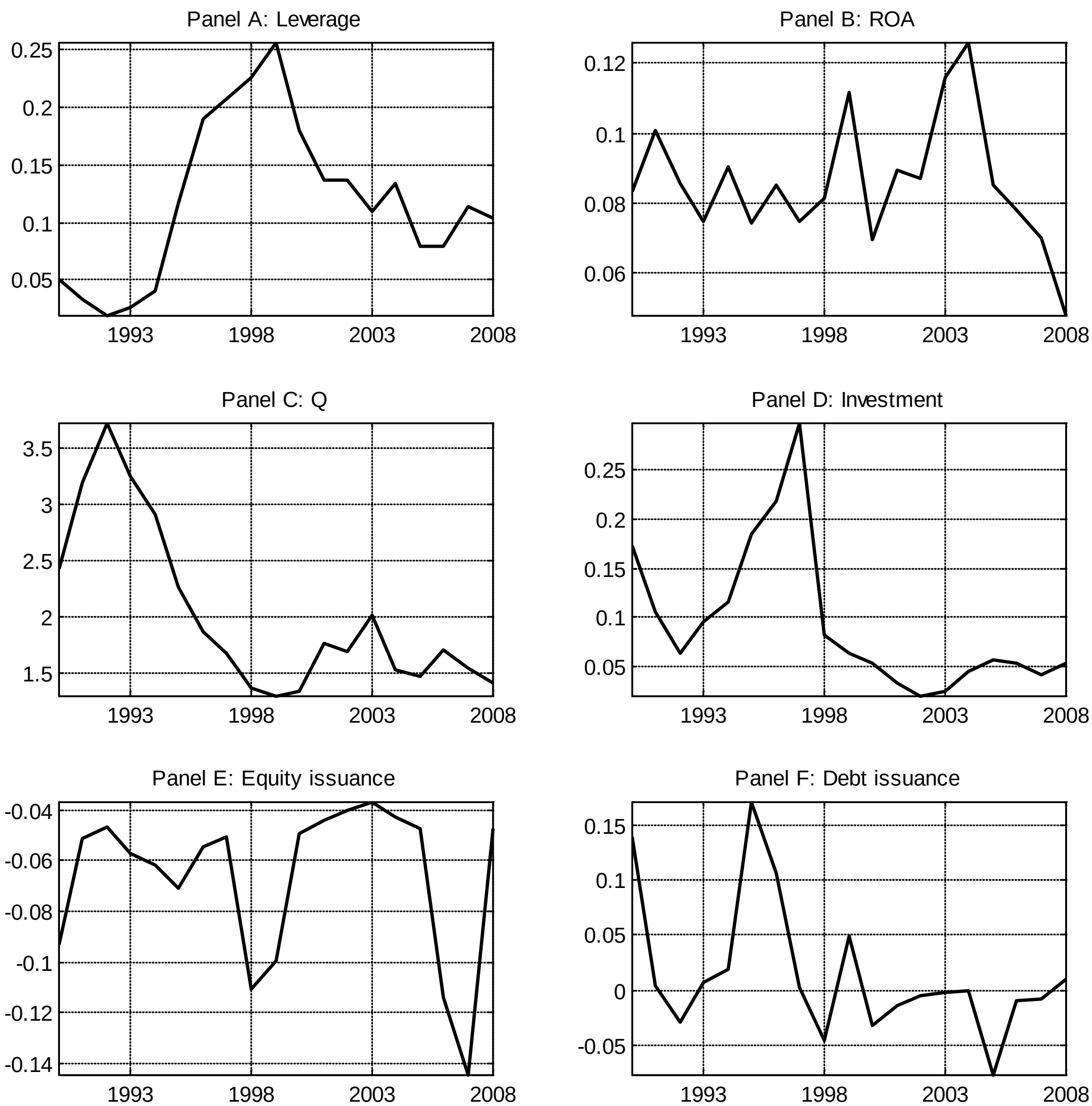
Panel A: Leverage
0.25
0.2
0.15
0.1
0.05
1993
1998
2003
2008
Panel B: ROA
0.12
0.1
0.08
0.06
1993
1998
2003
2008
Panel C: Q
3.5
3
2.5
2
1.5
1993
1998
2003
2008
Panel D: Investment
0.25
0.2
0.15
0.1
0.05
1993
1998
2003
2008
Panel E: Equity issuance
-0.04
-0.06
-0.08
-0.1
-0.12
-0.14
1993
1998
2003
2008
Panel F: Debt issuance
0.15
0.1
0.05
0
-0.05
1993
1998
2003
2008

Figure 2

Illustration of the Bayes factor's trade-off between model parsimony and fit. The horizontal axis presents possible realizations of the data, Y, and the likelihood of Y is on the vertical axis.

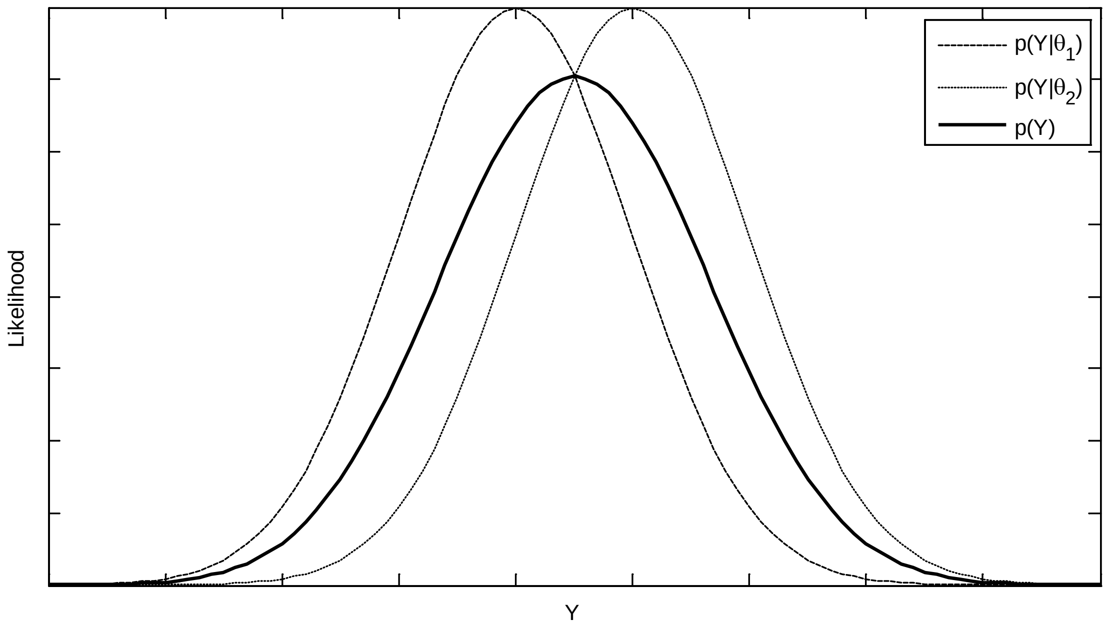

Figure 3

Illustration of the intuition behind the sequential updating of the particle filter. Dependence on model parameters is suppressed for ease of exposition.

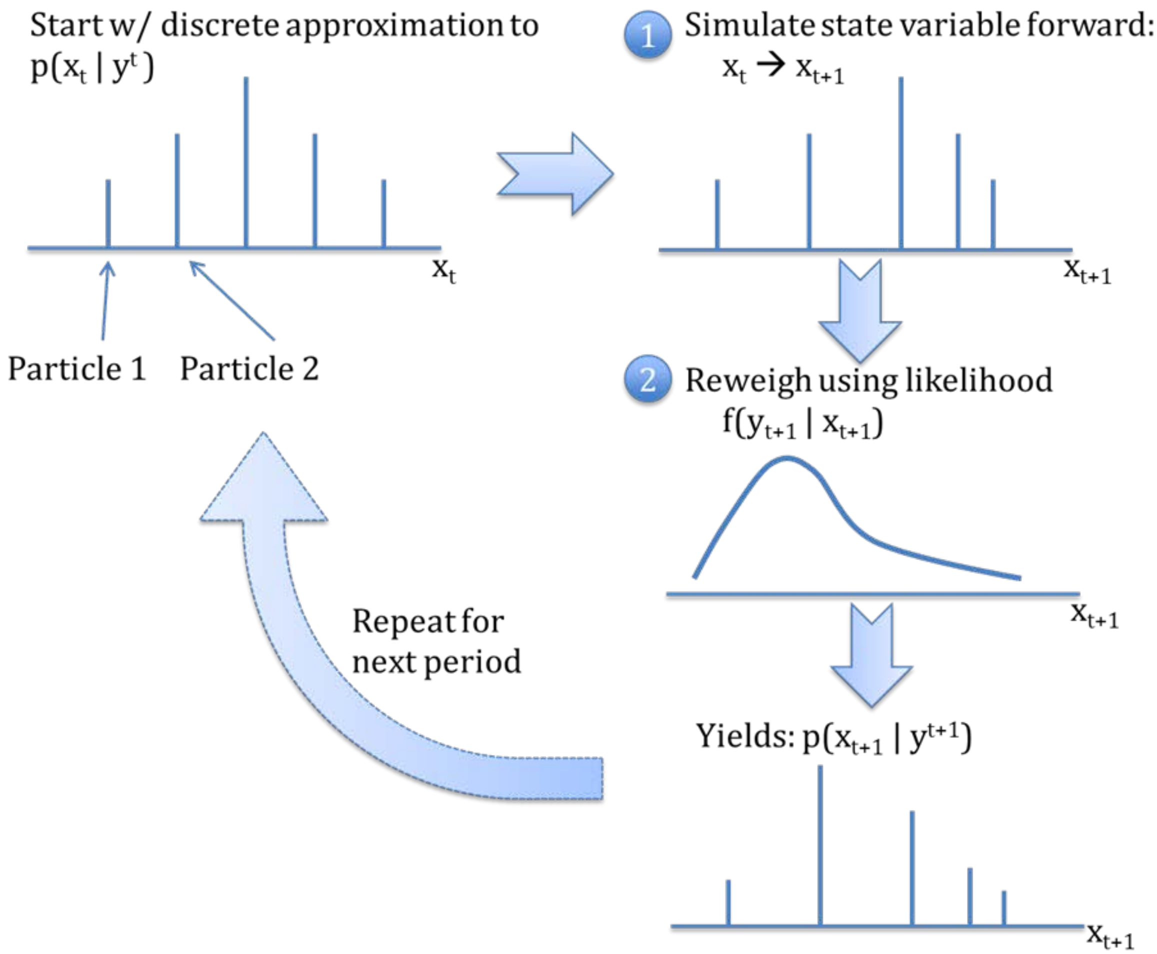

Figure 4

Plot of the relative model fit to leverage. Leverage is measured as the book value of debt net of cash divided by the book value of debt net of cash plus the market value of equity. Models are fitted to the five firms closest to median book assets in Compustat, over the period 1989-2008. The top plot shows the each individual year's log-likelihood for the random financing, pecking order and Goldstein, Ju and Leland (2001) models. The middle plot shows the cumulative log-likelihood and the bottom plot shows the Bayesian Information Criterion.

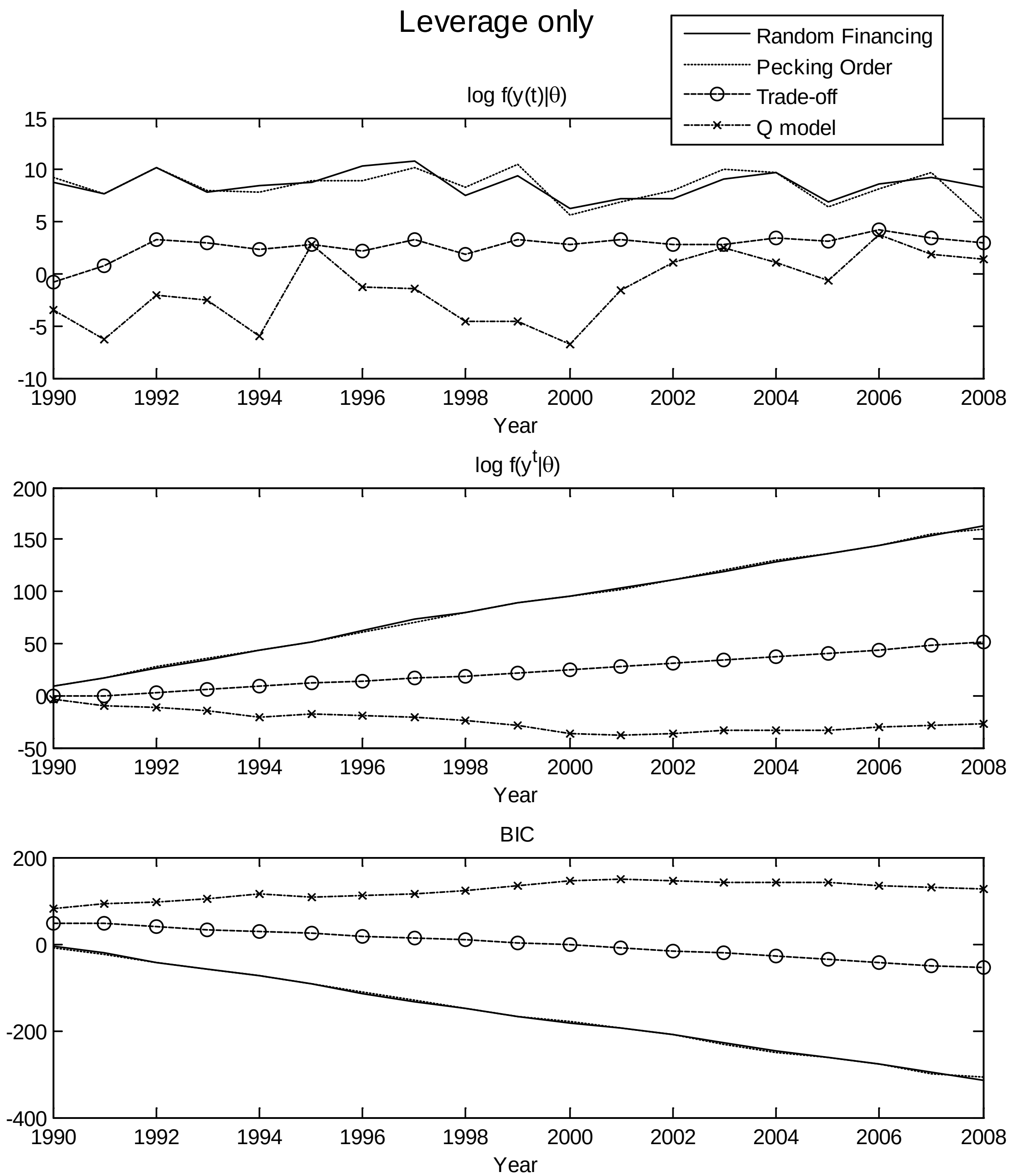

Figure 5

Plot of the relative model joint fit to leverage and return on assets (ROA). Leverage is measured as the book value of debt net of cash divided by the book value of debt plus the market value of equity. ROA is the operating income before depreciation divided by the book value of debt plus the market value of equity. For the random financing and pecking order models, ROA is assumed to be distributed iid Normal with the empirical mean and variance. Models are fitted to the five firms closest to median book assets in Compustat, over the period 1989-2008. The top plot shows the each individual year's log-likelihood for the random financing, pecking order and Goldstein, Ju and Leland (2001) models. The middle plot shows the cumulative log-likelihood and the bottom plot shows the Bayesian Information Criterion.

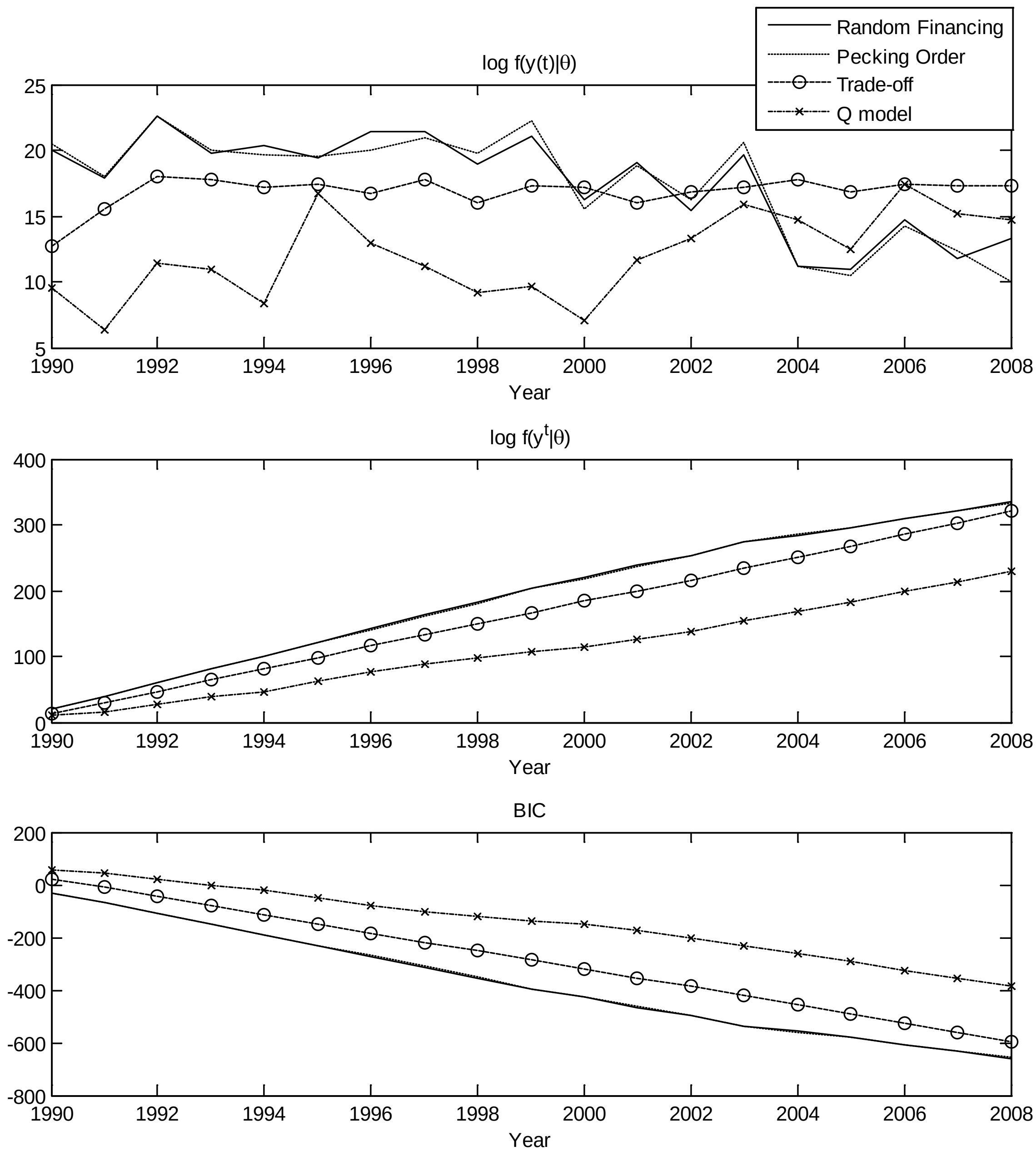

Figure 6

Model fit to observed leverage for the five firms closest to median total assets.

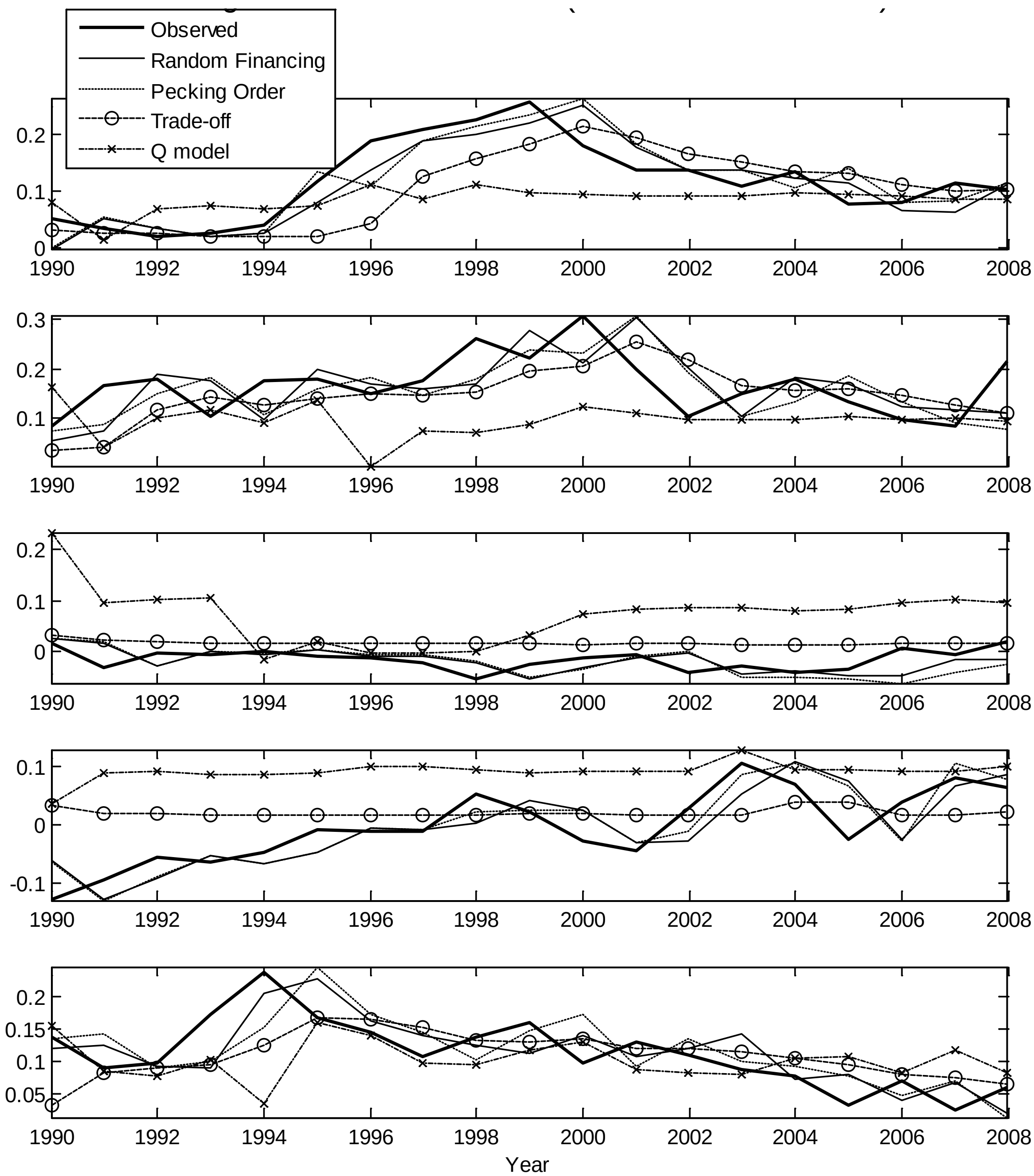

Figure 7

In-sample model fit to observed leverage for the five firms closest to median total assets. "In-sample" means that the Trade-off and Q models use the posterior mean of the state variable instead of the predictive mean.

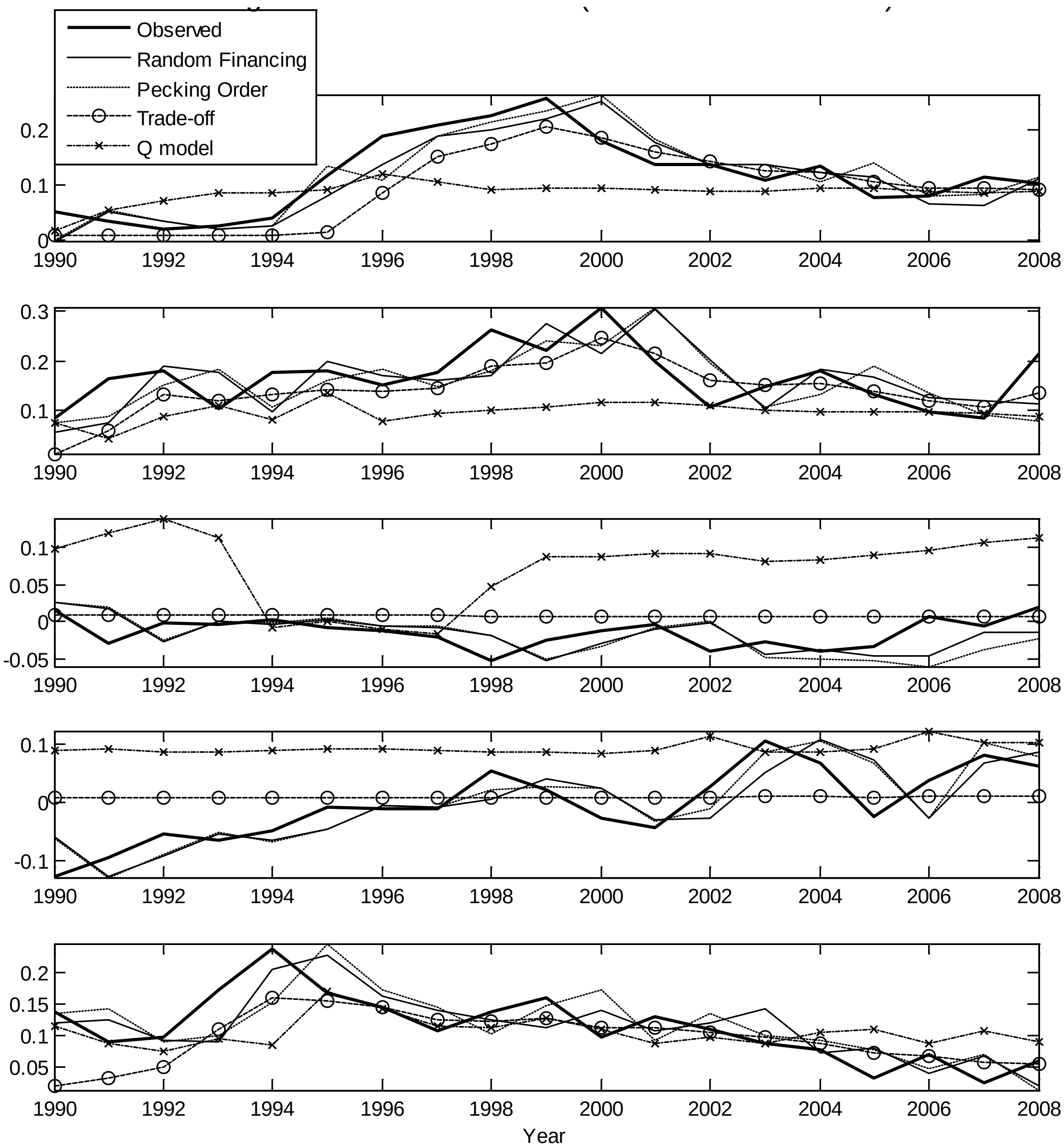

Figure 8

Model fit to observed return on assets (ROA) for the five firms closest to median total assets.

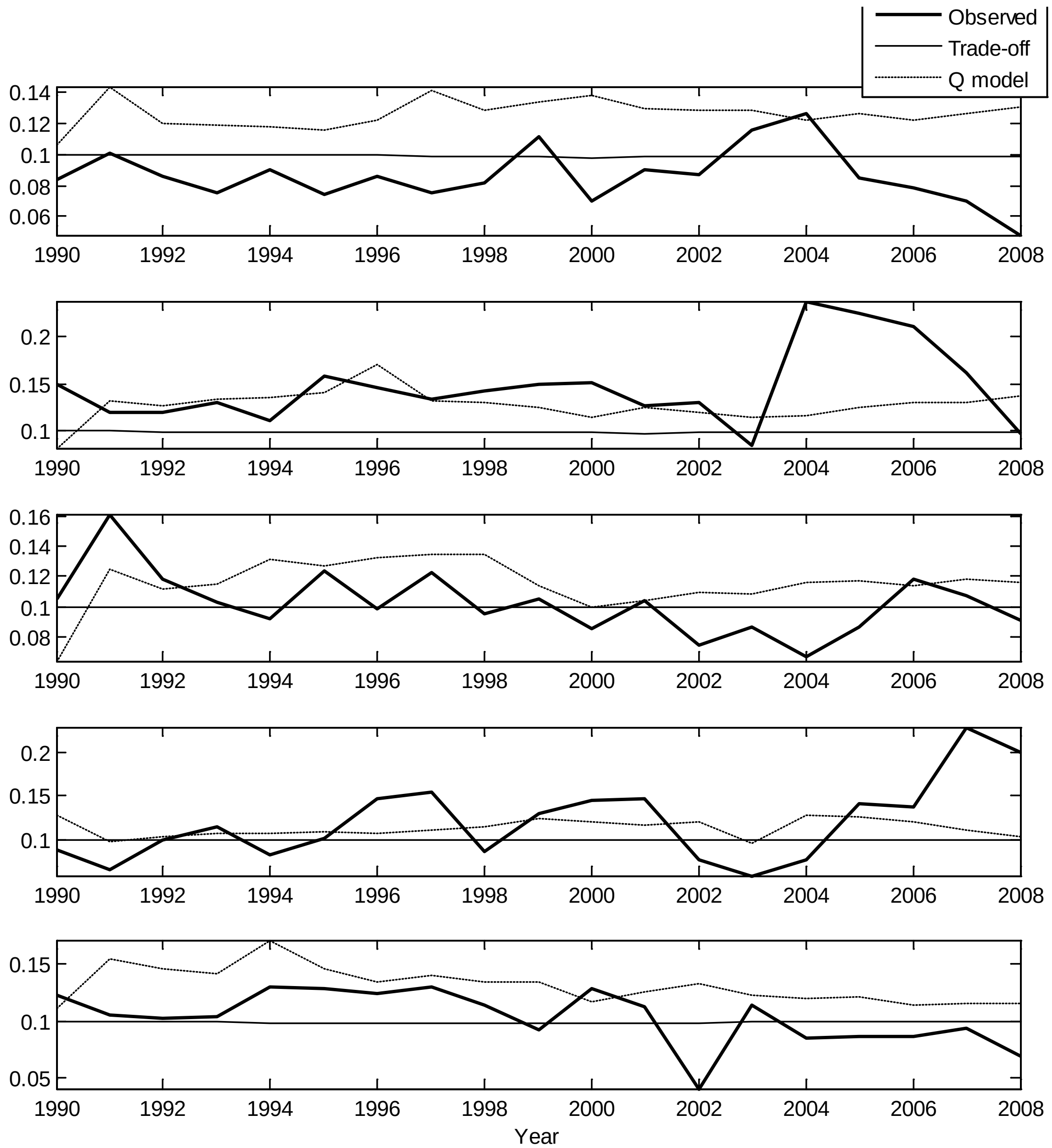

Figure 9

Model fit to observed investment (relative to PPE) for the five firms closest to median total assets.

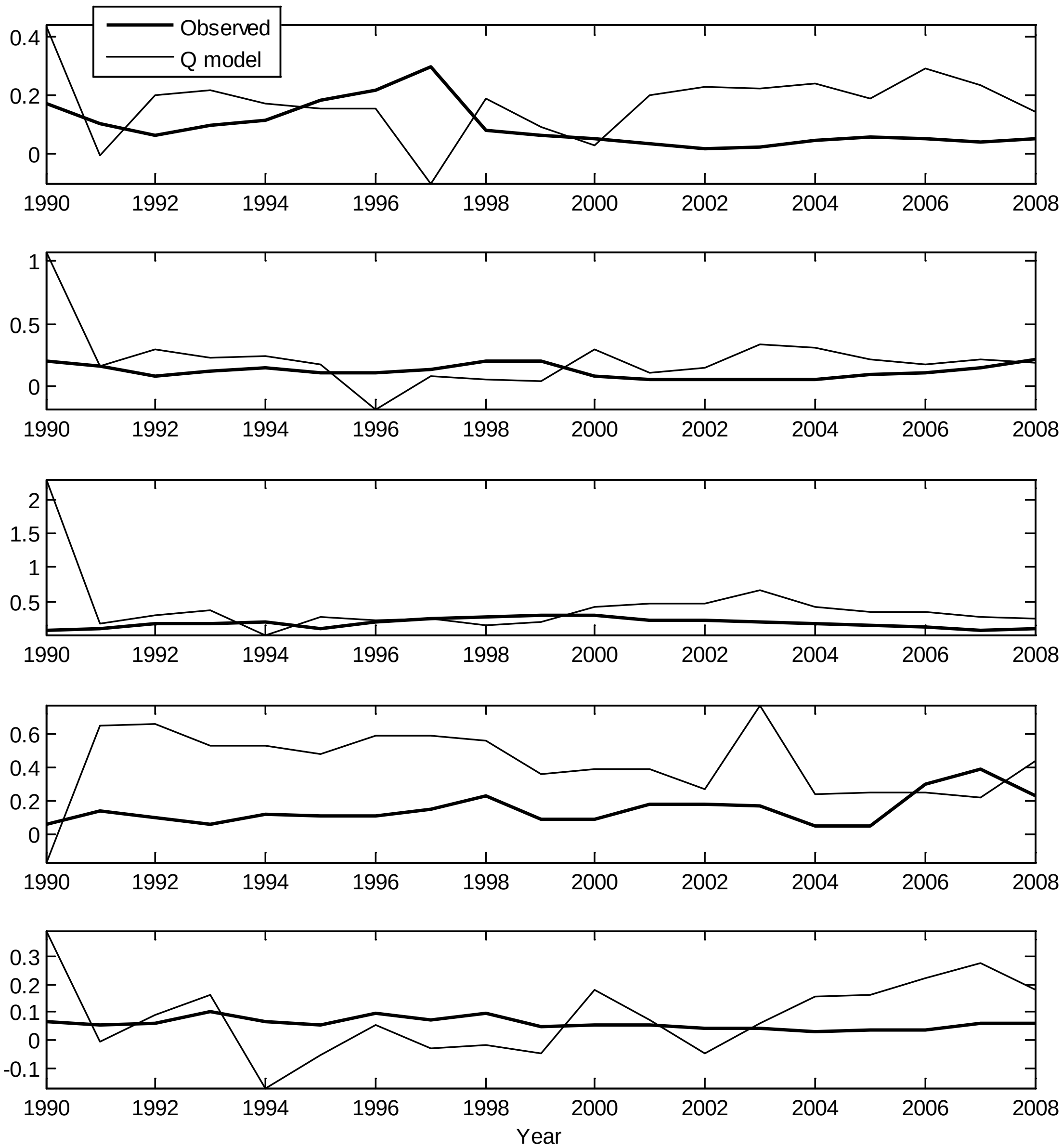

Figure 10

The top graph plots the delta, the state variable in the Goldstein, Ju and Leland (2001) against the model-implied market value of equity. The bottom plot shows the model-implied ROA (delta divided by debt plus market equity) as a function of the state variable.