European Management Review, Vol. ••, ••–•• (2014)
DOI: 10.1111/emre.12023

A Structural Approach to Handling Endogeneity in Strategic Management: The Case of RBV

ANUP MENON NANDIALATH[1], JEFFREY P. DOTSON[2] and RODOLPHE DURAND[3]

[1]Zayed University, Dubai, United Arab Emirates
[2]Brigham Young University, Provo, UT, USA
[3]HEC Paris, Jouy en Josas, France

In this paper we posit that the lack of consensus about empirical tests of resource based view (RBV) could be the result of endogenous resource picking on the part of firms. If resources are endogenously selected, regression based methods that examine their connection to firm performance will be mis-estimated. We show that traditional remedies for endogeneity do not resolve this problem when returns to resources are heterogeneous (as theorized under RBV) and when managers act with at least partial knowledge of the expected, idiosyncratic return (as theorized under the strategic factor market hypotheses). As such, we develop a Bayesian approach that solves this endogeneity problem by directly incorporating resource picking into the modeling framework. We illustrate the validity of our approach through the use of a comprehensive simulation study and show that our proposed approach outperforms traditional linear models (including traditional cures of endogeneity and unobserved heterogeneity) under a variety of conditions. Our findings suggest that: (1) research in strategy requires a more careful and deeper understanding of potential sources of endogeneity and (2) the use of Bayesian methods in management can help develop more theoretically motivated empirical approaches to hypothesis testing.

Keywords: resource based view; Bayesian modeling; endogeneity; structural modeling; competitive strategy

Introduction

The resource based view (RBV) represents a theoretically well accepted framework of rent creation. However, empirical evidence in support of the RBV is mixed (Barney and Arikan, 2001; Arend, 2006; Newbert, 2007; Crook *et al*., 2008), warranting further research in this area. In this paper, we suggest that the lack of empirical consensus may be the result of endogeneity inherent in the RBV. The strategic factor market (SFM) thesis, for example, stresses that some managers have superior expectations regarding a resource's rent-generating potential. Hence they may have greater skill in acquiring such resources. If managerial choices are driven by expectations regarding how their actions will influence firm performance, then such actions are endogenous by nature and must be accounted for in model estimation. Failure to do so can lead to serious mis-estimation of the relationship in question.

Correspondence: Anup Nandialath, Zayed University, Academic City, Al Ruwayyah, Dubai, UAE, E-mail: anup.nandialath@zu.ac.ae

We propose a unified model for treating endogeneity which closely mirrors the full information approaches suggested in the literature (e.g., Dotson and Allenby, 2010). Our choice is motivated by the importance of identifying the source of endogeneity within the RBV. While most empirical research treats endogeneity as a nuisance which needs to be corrected for, we argue that endogeneity is inherent within the RBV and hence has important implications for empirical designs. This is especially so as the effectiveness of traditional IV/treatment effect (e.g., selection models, matching, etc.) approaches, which are commonly employed in strategic management research (e.g., Shaver, 1998; Hamilton and Nickerson, 2003; Bascle, 2008; Semadeni *et al*., 2014), are dependent on two critical assumptions. The first assumption is that returns to resources are homogenous. Adherence to this condition is tenuous as it requires a violation of the central thesis of the RBV.

The second assumption is that managers and firms choose resources as if they did not have any knowledge of their potential return. The implications are further complicated when the element of evolution of resources in introduced. If firm resources evolve over time, the overall benefits of a resource may depend on existing/future complementarities with other resources, which would suggest that (1) returns to the same resource may be different between firms on account of variation in complementary resources and (2) managers may indeed foresee such potential complementarities in seeking appropriate resource configurations (see Argyres and Zenger, 2012). Violation of these two assumptions on heterogeneity and resource picking leads to a form of endogeneity referred to in the program evaluation literature as models with "essential heterogeneity" or models with "correlated random coefficients" (e.g., Heckman and Vytlacil, 2005). The presence of this type of endogeneity renders traditional methods to correct for endogeneity ineffective (see, Heckman *et al*., 2006).

A full information approach which addresses the concerns above proceeds by directly incorporating the managerial choice problem within the estimation framework. This approach gives us flexibility in imposing constraints on the optimization process that are based on theory. A significant example is the work by Arend and Levesque (2010), where they show that traditional statistical tests fail to capture the relationship between resources and performance when the choice behavior is endogenously modeled.

We illustrate the costs and benefits of both classical methods and our proposed Bayesian approach through the use of a series of simulation experiments. Simulated data enable researchers to better understand when and under what conditions an estimation approach can recover the true parameters used to generate the data. As a result, it provides researchers with a formal way to assess the performance of various estimation methods given a particular data structure of interest to the researcher (ordinary least squares (OLS) and instrumental variable (IV) regressions in our case).

The use of simulation to study the theoretical properties of estimators has a long-standing tradition in a variety of fields including strategic management (Certo and Semadeni, 2006; Arend and Levesque, 2010; Henderson *et al*., 2012; Semadeni *et al*., 2014; Shaver, 2007), finance (Prabhala, 1997; Chang and Dasgupta, 2009) and marketing (Fornell and Larcker, 1981; Swait and Louviere, 1993), and is particularly appropriate in this context as our goal is not to prove or disprove the RBV, but rather to help disentangle the different sources of endogeneity and explain the existence of mixed support in the existing body of empirical literature. As we know the value of parameters used for simulating the data, we can accurately test which estimation methods are more capable than others at recovering the "true" associations between factors (say resources and performance).

Results from our simulation studies show that in the presence of endogenous resource picking, our proposed model is able to recover the true relationship between resource and performance, outperforming both traditional OLS regressions and an IV counterpart. We find that the extent of bias in traditional OLS and IV regressions can be quite significant, even resulting in sign reversals for relationships that are (in simulation) positive, like the association between a resource and firm performance. We also show that our model is robust to modeling assumptions in the decision making process. For example, if resources are exogenous to the system of study, both classical methods and our proposed model will effectively recover the true relationship. Furthermore, recent empirical evidence suggests how difficult it is to exclude luck as a factor related to superior performance (Henderson *et al*., 2012). Models that do not account for luck as a possible explanation are incomplete and may lead to faulty inference. Our approach lets us measure the magnitude of error in a manager's resource picking ability, thus allowing us to formally test if superior performance realized by a firm is the result of managerial skill or other exogenous factors such as luck (Barney, 1986) that is, when the resource picking error size is small (skill) vs. large (exogenous factors/luck).

Testing the RBV

Within the RBV framework a firm is treated as a bundle of resources and capabilities where there is significant cross-firm heterogeneity in resources. The RBV suggests that firms that have access to specific types of resources will generate superior performance relative to its competitors. It is therefore critical to identify the specific types of resources that can lead to such performance differentials. According to Barney (1991) and followed by many subsequent works, resources that possess specific properties (e.g., valuable, costly to imitate, rare and non-substitutable), and are heterogeneously distributed across firms, will generate superior competitive advantages.

Over the years, authors have questioned the underlying assumptions of the RBV. For instance, it has been argued that the RBV is tautological by nature (Priem and Butler, 2001). It has also been suggested that there has been an overemphasis on factor markets relative to demand side factors, thus potentially hampering external validity (Foss and Hallberg, 2010). Powell (2001) introduces the idea that firms with special resources (i.e., rare and valuable) may very well be below-average performers, while firms with no special resources can over-perform the market. This gives rise to the argument that mere possession of resources is insufficient, but that competitive advantage may reside in the ability of the managers to exploit the value of those resources (Penrose, 1959) or in the capacity of the organization itself (Durand, 2002).

Despite compelling potential for the RBV to explain superior firm performance, researchers have been skeptical of the empirical tests that are used in practice (Collis, 1994; Godfrey and Hill, 1995; Priem and Butler, 2001). Several questions come to the forefront. First, following the Penrosian tradition, are resource properties objective or subjective? Furthermore, the inherent properties of resources such as rarity and inimitability are often unobservable, thus making it difficult for researchers to identify and measure a "resource" with any degree of confidence (Godfrey and Hill, 1995; Arend and Levesque, 2010).

Even if one assumes that the resources have been identified and measured, the standard approach used to test the RBV has focused on regression models that examine the effect of resources on performance (e.g., Henderson and Cockburn, 1994; Miller and Shamsie, 1996; Mowery *et al.*, 1996).[1] While resources are tangible, such as capital, they tend not to offer sustainable advantages simply because they do not satisfy the properties of rarity or inimitability. Recognizing this, other studies have focused on intangible assets that are rarer and harder to imitate but difficult to observe with high validity and reliability (e.g., Chatterjee and Wernerfelt, 1991; Bierly and Chakrabarti, 1996; Szulanski, 1996).

Apart from measurement issues, the RBV is also viewed as a firm-specific theory. While most traditional, regression based methods answer the question on how resources affect the performance of the average firm, heterogeneity in firm performance is often ignored. A major contribution of Hansen *et al.* (2004) and other works inspired by Bayesian methods (Hahn and Doh, 2006; Kruschke *et al.*, 2012) is their use of a hierarchical Bayesian regression model that allows researchers to study firm specific outcomes and, by extension, the distribution of outcomes taken across firms. This is an important issue as the prior literature suggests that average effects may not be informative enough to lend credence to the RBV, which pays heed to positive outliers (e.g., Wiggins and Ruefli, 2002).

Problem of endogeneity and traditional solutions

The role of endogeneity and the determination problems it entails have been well acknowledged in the strategic management literature (e.g., Shaver, 1998; Hamilton and Nickerson, 2003; Semadeni *et al.*, 2014; Bascle, 2008 among others). Hamilton and Nickerson (2003) provide a thorough review on methods available to "correct for endogeneity." However, they do not delve into the underlying sources of endogeneity, which may play an important role in testing theories such as the RBV. Identifying the source of endogeneity is important as methods used to correct for them depend on crucial assumptions on the nature of endogeneity (Heckman and Vytlacil, 2005).

In the field of strategy in particular and management in general, endogeneity is typically treated as an omitted variables problem (among the other sources are errors-in-variables, autoregression and simultaneous causality). For instance, from the SFM hypothesis a critical variable which influences firm performance is the ability of managers, which may not be observable to the researcher. Hence the unobserved ability is driven into the error term of the regression model thereby generating a correlation between the variable of interest (the resource) and the error term. This violates the textbook conditions required for generating an unbiased estimate of the relationship between the resource and performance, thereby tainting inference.

To correct for this problem, researchers may use one of several approaches. If the endogenous variable is continuous then using two-stage least squares with appropriate instrumental variables (IV) may help correct for the endogeneity bias. If longitudinal data is available, the use of fixed/random effects methods may be warranted. If the endogenous variable is binary (such as observing some form of choice behavior), the use of self-selection models (Heckman, 1979) or matching methods (Rosenbaum and Rubin, 1983) is recommended.[2] However, the choice of method tends to be driven by the type of data available (continuous vs. binary variable), but rarely do we find a discussion on the assumptions needed to ensure that endogeneity is properly handled given the context of the problem being studied. For instance, fixed effects, a very popular method in the strategy literature, controls for endogeneity only under the assumption that the unobserved heterogeneity is time invariant. Typical hierarchical Bayes models (such as the one used by Hansen *et al.*, 2004) are conceptually equivalent to random effect models. These models also require time invariance properties for heterogeneity to generate valid results. In our illustrative example of managerial skill, where ability constitutes the unobserved variable, this implies that there is no scope for learning and abilities remain constant over long time periods. Given the substantial amount of literature on learning and cognition (Szulanski, 1999; Zollo and Winter, 2002; Coff, 2010), this would seem to be a rather untenable assumption. Other commonly used methods to correct for endogeneity such as dynamic panel models (Blundell and Bond, 1998) and selection/treatment effect models (with

[1]It should be noted that apart from regression based large sample studies, researchers have also followed alternative paths such as using the case study approach (e.g., Kotha, 1995) or survey based designs (e.g. Newbert, 2008) to validate the RBV. These approaches while offering an alternative empirical design could still suffer from endogeneity concerns.

[2]In this paper we do not discuss aspects of treatment effect models such as Heckman's selection approach or matching methods. These models work when the treatment is binary but less is known when the endogenous variable has a continuous support.

the exception of matching estimators[3]) require high quality instrumental variables (exclusion restrictions) to ensure unbiased inference. However the validity of such instruments can be highly questionable under many circumstances (Semadeni *et al*., 2014). While there exist a number of tests to check for the validity of instruments (e.g., test of overidentifying restrictions), the effectiveness of these tests to detect high quality instruments can be questionable (Whited and Roberts, 2012). Therefore, it might be difficult for researchers to come up with rationally justifiable exogenous instruments (e.g., Conley *et al*., 2012).

While the above arguments are necessarily statistical, an important aspect which often does not find its way into discussions rests with the underlying nature of the endogeneity. Inference using traditional methods of correcting for endogeneity rests on two key assumptions regarding the sensitivity of resource to performance (the β in a regression of resources (X) to performance (Y)), namely:

1. the effect of the resource on performance is homogenous, implying that all firms receive the same marginal benefit from the resource; and
2. agents (managers) are assumed to have no knowledge about β of the resource in influencing performance or even if they do, such information is not used by them in their resource picking decision.

Both these assumptions are highly questionable when testing the RBV. Assumption 1 requires lack of heterogeneity in the sensitivity of the resource to performance across firms, which is counterintuitive to RBV theory as by definition, returns to resources are assumed to be heterogeneous across firms. Second, ignoring the knowledge of the resource performance sensitivity (even in expectation) seems counterintuitive in most circumstances as it is extremely difficult to envision that managers do not use their skills in identifying and acquiring resources. The literature terms this type of endogeneity as "essential heterogeneity" (Heckman and Vytlacil, 2005) or "correlated random coefficients" (Heckman and Vytlacil, 1998). In the presence of "essential heterogeneity," it has been shown that traditional cures for endogeneity such as IV based approaches can be problematic (Heckman and Robb, 1985; Heckman and Vytlacil, 2005; Heckman *et al*., 2006). To stress the point, it has also been suggested in the literature that "the cure (IV based methods) can be worse than the disease (endogeneity in OLS regression)" (Heckman and Vytlacil, 2005: 271).[4]

The problem of inference

The above discussion suggests that a robust test of the RBV should effectively deal with the problem of essential heterogeneity. Recall that as a fundamental goal, empirical tests of the RBV are designed to determine if resources have a positive, causal impact on firm performance. Nine out of ten empirical studies follow this line of reasoning, according to Newbert (2007). To illustrate, let Y_i represent observed performance for firm i. Let r_i represent the resource configuration and C_i denote the set of relevant control variables. In this setting the classical test of the RBV is accomplished by first estimating the following cross-sectional regression:

$$Y_i = \alpha + \beta r_i + \gamma C_i + \eta_i + \varepsilon_i$$

Consistent with our example, η_i is a measure of unobserved (to the researcher) managerial ability;[5] ε_i is the error term and α, β *and* γ are parameters to be estimated. Validity of the RBV is established by evaluating if the partial derivative of performance with respect to the resource as given by $\frac{\partial E(Y)}{\partial r}$ is positive. In the linear model presented above, this occurs when $\beta > 0$. This interpretation is only valid under certain conditions. One such condition for valid inference on β rests in the assumption that $\eta_i = 0$ which implies that there are no omitted correlated variables affecting performance. This assumption is more tenuous to impose in practice. If as the SFM thesis suggests managers indeed use their resource picking skills, ensuring that $\eta_i \neq 0$ and $Cov(r_i, \eta_i) \neq 0$, implies that corrections for endogeneity is required in order to obtain reliable inference for β.

Traditional approaches to correcting for endogeneity depends on the kind of data available to researcher, e.g. cross-sectional data or longitudinal.

Cross sectional data. If $\eta_i \neq 0$ and it is unobserved the resultant regression model looks as follows

$$Y_i = \alpha + \beta r_i + \gamma C_i + \overline{\varepsilon}_i,$$

[3]While matching estimators do not require instruments, they only identify treatment effects based on a strong and untestable assumption that selection into treatment occurs on observables only.

[4]Note that other approaches such as the local average treatment effect (LATE) have been suggested to deal with heterogeneous responses when the treatment variable is binary (see Imbens and Angrist, 1994). Since inference is restricted to a particular sub-population and identification of the sub-population is not clear coupled with the binary treatment variable, makes this approach less interesting for the purpose of testing the RBV.

[5]The assumptions on what is unobserved can be problem specific and has no bearing on the results of our analysis.

where $\overline{\varepsilon}_i = \varepsilon_i + \eta_i$. Thus the unobserved managerial ability is now absorbed into the error term of the regression. Since $Cov(r_i, \eta_i) \neq 0$, it implies that inference on β is biased on account of this endogeneity problem. A solution for this problem rests in identifying an instrument Z which shares the following properties

$$Cov(Z, r) \neq 0$$

and

$$Cov(Z, \overline{\varepsilon}) = 0.$$

Under these conditions, traditional IV estimators guide precise inference on the coefficient of resource on performance β in large enough samples

$$p\lim \hat{\beta}_{IV} = \frac{Cov(Z, Y)}{Cov(Z, r)} = \beta.$$

This approach (using IV methods) is fundamental to many techniques used to correct for endogeneity. The most popular techniques also used in the strategic management literature include two stage least squares and selection models (in the spirit of Heckman, 1979; Heckman and Robb, 1985). This modeling approach works well in the presence of a good instrument which shares the properties mentioned above. However, finding such instruments is difficult in practice (e.g., Conley *et al*., 2012).

Longitudinal data. When good instruments are tough to find, researchers can at times rely on time series properties of the data to control for endogeneity. One such approach is to run fixed effect regressions. To illustrate this point we go back to our original regression model with an additional subscript capturing the time series element on all variables, with the exception of η as the fixed effects estimator requires the unobserved variable to be time invariant:

$$Y_{it} = \alpha + \beta r_{it} + \gamma C_{it} + \eta_i + \varepsilon_{it}.$$

A fixed effects estimator works by subtracting each component of the above model from their respective means. Thus, the revised regression equation is as follows:

$$\begin{aligned} Y_{it} - \overline{Y}_i = {} & (\alpha - \overline{\alpha}) + \beta(r_{it} - \overline{r}_i) + \gamma(C_{it} - \overline{C}_i) \\ & + (\eta_i - \overline{\eta}_i) + (\varepsilon_{it} - \overline{\varepsilon}_{it}). \end{aligned}$$

Given the time invariance property, variables without the t subscript vanish from the model above since the time invariance implies that the means are identical to the values within each firm. The reduced form model which is estimated then looks like:

$$\ddot{Y}_{it} = \beta \ddot{r}_{it} + \gamma \ddot{C}_{it} + \ddot{\varepsilon}_{it}.$$

Valid inference on β can now proceed as unobserved ability, which is assumed to be constant over time has been eliminated from the system. Therefore a simple demeaning of the data can eliminate endogeneity provided the assumptions behind the source of the endogeneity can be justified.

To address more complicated dependence structures where unobserved heterogeneity is combined with another form of endogeneity called simultaneity, dynamic panel data methods have been proposed. The most popular approach to estimating dynamic panel data methods is using lagged values of independent variables as instruments and estimation using generalized method of moments (GMM) (in the spirit of Arrelano and Bond, 1991; Blundell and Bond, 1998; among others). However, the assumptions on the nature of endogeneity behind these models mirror those of the IV and fixed effects estimation and hence the method suffers from the same critiques when it comes to solving endogeneity on account of essential heterogeneity.

Essential heterogeneity. While the solutions for both cross sectional data and longitudinal data depend on instrument exogeneity and time invariance of fixed effects, a deeper problem arises when (1) the effect of interest as captured by β is heterogeneous and (2) if managers have private information regarding the efficacy of the resource to performance and if they use that in their resource acquisition decision. Under these conditions $Cov(\beta, r_i) \neq 0$ and this complicates the endogeneity problem. This is referred to in the literature as models with “essential heterogeneity” (Heckman *et al*., 2006). Heckman *et al*. (2006) shows that when the resource sensitivity is correlated with the resource itself, standard instrumental variables fail to address the endogeneity issue.

To illustrate this point we highlight the simple proof from Heckman *et al*. (2006). Suppose $\beta = \overline{\beta} + \eta$ where $\overline{\beta}$ captures the average effect of the resource on performance and η captures the superior knowledge or skill that a particular manager possesses. The original performance equation to be estimated can now be written as:

$$Y_i = \alpha + \overline{\beta} r_i + \gamma C_i + (\eta r_i + \varepsilon_i).$$

Assuming an instrument Z is available, in the standard IV setup, while Z might be uncorrelated with ε *and* η, it also requires Z to be uncorrelated with $\eta r_i + \varepsilon$. This can only occur if $Cov(\eta, r_i) = 0$; that is, when managers do not use any private information or skill in their resource acquisition decision, which implies that resource picking is indeed exogenous and it is counter-intuitive to the basic ideas behind the SFM thesis. However, if managers use their private information, then standard IV estimation does not recover the true effect of resource on performance (Heckman and Robb, 1985; Heckman *et al*., 2006). To make matters worse, using IV based corrections for endogeneity might actually lead to more bias than using simple OLS. Paraphrasing from Heckman and Vytlacil (2005: 271) "The cure may be worse than the disease."

Solutions to the problem

Various solutions have been proposed when dealing with heterogeneous responses. The first approach to dealing with the problem focuses on estimating the marginal treatment effect which still depends on the availability of a good instrument (e.g., Heckman *et al*., 2006). The second approach uses the statistical properties of data available to generate instruments from within such as higher order moments estimators (e.g., Erickson and Whited, 2002) or identification through the presence of heteroskedasticity (e.g., Rigobon, 2003) or opt for a completely instrument free approach as in latent instrumental variable (LIV) approaches (e.g., Park and Gupta, 2012). However these methods rely less on the theoretical structure underlying the decision problem and depend on statistical properties of the data. Therefore, it might lead to conclusions often at odds with theory which is not desirable. A third approach that has been proposed to deal with endogeneity focuses on building structural models which integrate theory with empirical models and is also known as the full information approach to dealing with endogeneity (e.g., Yang, Chen and Allenby, 2003; Dotson and Allenby, 2010; Otter *et al*., 2011 among others).

In this paper, we propose a variant of the third approach and suggest that modeling of the endogeneity (manager's decision process) directly in the empirical design provides an ideal tool to test the RBV. We choose the third approach as it provides us with several advantages. First, unlike the first approach, the full information approach does not depend on the availability of a valid instrumental variable. Second, unlike the second approach discussed above, the solution to endogeneity is not atheoretical. On the contrary the full information approach can effectively handle "essential endogeneity" by modeling the decision process within the empirical estimation model. Next, we develop a model of firm performance where resources are endogenous to the system and the managerial decision process (which captures endogeneity) is directly modeled through a supply side system of equations.

A model of resource selection and firm performance

In this section we begin by specifying a general model of resource picking/managing and firm performance and discuss a corresponding Bayesian estimation strategy. We then show how this model can be modified to accommodate a variety of data structures commonly encountered in the field of strategy. A simple illustration (e.g., one endogenous resource and one control variable) of this general approach appears in the development of our simulation study. It is our hope that this simple model will help less technical readers develop an intuition for our approach (and its corresponding benefits). We include the derivation of the general model in the paper for more technical readers and for completeness.

We begin by following Dotson and Allenby (2010) and Otter *et al*. (2011) and specify a structural model to capture the relationship between resources and firm performance. Firm performance is governed by the functional form presented in equation (1):

$$y_{it} = \beta_{0i}\left(\prod_{k=1}^{K} x_{kit}^{\beta_{ki}}\right)e^{\varepsilon_{it}}. \tag{1}$$

where y_{it} is a performance outcome realized by unit i in time t and $\{x_{kit}\}$ denotes the set of potentially endogenous inputs which impact performance. The log-linear version of equation (1) directly translates into the classical linear regression framework widely used in empirical studies in strategic management. Resource picking is incorporated into the model by assuming that managers pick the inputs $\{x_{kit}\}$, in order to maximize the performance outcome presented in equation (2):

$$\begin{gathered}\max_{\{x_{kit}\}} \sum_i \sum_t \left(\beta_{0i}\left(\prod_{k=1}^{K} x_{kit}^{\beta_{ki}}\right) - \sum_{k=1}^{K} p_{kit}x_{kit}\right), \\ \textit{subject to} \sum_i \sum_t x_{kit} \le M_k\end{gathered} \tag{2}$$

where p_k denotes the cost and M_k denotes a boundary constraint for each input, k. For instance, suppose the performance outcome measured by y_{it} is revenue and p_k is measured as a monetary cost, then equation (2) represents the constrained optimization of a profit function for resources. Profit maximization, however, is just a special case of the more general maximization problem faced by firms. The structure of the model allows for any performance outcome to be modeled. This approach is a simplified version of the simulation model used by Arend and Levesque (2010).

Specifically, in this setup, firms attempt to maximize performance outcomes subject to the constraint that they are able to access a sufficient level of valuable resources. The inclusion of a resource acquisition model (equation 2) allows us to formally incorporate the managerial challenge of acquiring resources directly into our empirical model.

Given this optimization problem, we can employ the method of Lagrange to solve for the benefit maximizing values of the inputs, $\{x_{kit}^*\}$:

$$\ln(\mathbf{x}_{it}^{*})=\begin{bmatrix}\ln(x_{1it}^{*})\\ \vdots\\ \ln(x_{Kit}^{*})\end{bmatrix}=\begin{bmatrix}(\beta_{1i}-1) & \cdots & \beta_{Ki}\\ \vdots & \ddots & \vdots\\ \beta_{1i} & \cdots & (\beta_{Ki}-1)\end{bmatrix}^{-1} \times\begin{bmatrix}\ln(\lambda_1+1)-\ln(\beta_{0i})-\ln(\beta_{1i})+\ln(p_{1it})\\ \vdots\\ \ln(\lambda_K+1)-\ln(\beta_{0i})-\ln(\beta_{Ki})+\ln(p_{Kit})\end{bmatrix}. \quad (3)$$

The solution to this constrained maximization problem forms the basis of our model for endogenous resource selection. If our performance outcome function is deterministic (or is stochastic with symmetric errors) and managers behave optimally, they will utilize levels of inputs equal to $\{x_{kit}^{*}\}$.

Deviation from optimality

Although it may seem like the assumptions associated with managerial performance maximizing behavior presented in equation (2) are extreme (i.e., *ex ante* perfect knowledge of the benefits that may accrue from the resource and cost structure), our model allows for deviations from optimality through the introduction of error into the system presented in equation (3). This involves introducing a structural error term to the resource-picking equation specified above. In the formulation described in equations (1) and (2), this is accomplished by assuming that managers have imperfect knowledge of the cost associated with acquiring a specific resource, or that managers pick resources using a cost of $p_{kit}^{*}=p_{kit}e^{\zeta_{it}}$, where ζ_{it} captures perceptual deviations from the actual cost of the resource. Introduction of the structural error term through cost follows directly from the discussion above.

In our model, optimal behavior taken on the part of managers suggests that they have superior skill in identifying the properties of a particular resource and hence pick only those resources that will maximize firm performance. On the other hand, when the realized deviations of ζ_{it} are large it could mean that managers have little information about the properties of the resource and therefore exhibit behavior that appears more random and generates outcomes that can be attributed to chance. Under the latter scenario, firms that achieve superior performance may do so because of luck (Denrell, 2004; Denrell and Fang, 2010). Alternatively, it could reflect conditions where managers have not identified the VRIO resource correctly due to a number of factors such as causal ambiguity, managerial inattention or other socio-political issues (Arend and Levesque, 2010). In our proposed framework we can formally test for the validity of this assumption. Our estimation process allows us to estimate the variance associated with the supply-side model. By examining the estimated variance of the resource-picking equation we can determine if it is skill or other exogenous factors that drives firm performance. If other exogenous factors are driving firm performance the corresponding estimated variance of the resource selection equation should be large and the estimated parameters of the structural model should be equivalent to those generated from descriptive models (e.g., OLS).

Estimation of shadow prices

A second benefit of our model is the inclusion of the allocative efficiency property, as manifest through the parameter λ_k, the Lagrange multiplier or shadow price of a given input.

λ_k is the slope of the line tangent to the net-benefit function at the optimal value of x_k. If utilization of input k is unconstrained, then λ_k will be equal to 0. If the estimated value of λ_k is greater than 0, it implies that the firm can increase performance by increasing the level of resource utilized, x_k. If the manager is optimally allocating resources across all inputs x_k then λ_k should be equal across all k (i.e., the marginal increase in performance associated with the relaxation of the resource constraint is equivalent across all inputs). Examination of λ_k informs us about the degree to which resources are being optimally utilized within the firm. This can allow us to directly study resource deployment, a necessary but insufficient condition for superior performance, that is, to represent what Durand (2002) called a capable organization or Tang and Liou (2010) a firm's configuration.

Likelihood and estimation

We estimate the proposed model in a Bayesian framework using Markov chain Monte Carlo (MCMC) methods. As such, we must first derive the likelihood for the observed data (both output y and input X). This can be accomplished using change-of-variable calculus and re-expressing the likelihood for the observed data in terms of realizations of the error distribution ($\hat{\varepsilon}$ and $\hat{\zeta}$).

$$\ell(data \mid else)=\prod_i\prod_t \pi(\hat{\varepsilon}_{it})\pi\left(\left\{\hat{\zeta}_{kit}\right\}\right)\left|J_{\substack{it\\ \zeta\to x}}\right|. \quad (4)$$

Where the quantities $\hat{\varepsilon}_{it}$ and $\left\{\hat{\zeta}_{kit}\right\}$ denote the residual performance and resource picking errors that can be computed from equations (1) and (3) and $\left|J_{\substack{it\\ \hat{\zeta}\to x}}\right|$ is the Jacobian term that captures dependencies in the mapping of $\hat{\zeta}\to x$.

$$\hat{\varepsilon}_{it}=\ln(y_{it})-\left(\ln(\beta_{0i})+\sum_{k=1}^{k}\beta_{ki}\ln(x_{kit})\right). \quad (5)$$

$$\begin{bmatrix} \hat{\zeta}_{1it} \\ \vdots \\ \hat{\zeta}_{Kit} \end{bmatrix} = \begin{bmatrix} (\beta_{1i}-1) & \cdots & \beta_{Ki} \\ \vdots & \ddots & \vdots \\ \beta_{1i} & \cdots & (\beta_{Ki}-1) \end{bmatrix} \ln(\mathbf{x}_{it}^{*}) - \begin{bmatrix} \ln(\lambda_1+1) - \ln(\beta_{0i}) - \ln(\beta_{1i}) + \ln(p_{1it}) \\ \vdots \\ \ln(\lambda_K+1) - \ln(\beta_{0i}) - \ln(\beta_{Ki}) + \ln(p_{Kit}) \end{bmatrix}. \quad (6)$$

$$\left| J_{\zeta \to x} \right| = \left| \sum_{k=1}^{K} \beta_k - 1 \right|. \quad (7)$$

Given the specification of the likelihood, Bayesian estimation proceeds by recursively generating draws from the full conditional distributions of all model parameters (Rossi *et al.*, 2005). The inclusion of the Jacobian term in equation (4) prevents us from utilizing conjugate results in order to implement a Gibbs sampler for model estimation. A hybrid sampler is used instead where a subset of the parameters are drawn using the Metropolis-Hastings algorithm (Chib and Greenberg, 1995). Details of the estimation algorithm for a simplified version of this model appear in the technical appendix.

Simulation description and results

In this section, we simplify the general model described above and demonstrate through a series of simulation studies that joint modeling of both performance and resource picking allows us to recover the true causal impact of resources on firm performance, that is, the value of the parameters used to simulate data. We then permute the simple model to accommodate a variety of data structures of interest to empirical researchers in the field of management/strategy, including panel data and serial autocorrelation.

Simulation description

To ease exposition we employ the following simplification of the general model by focusing only on a resource picking problem. Suppose we observe some measure of performance, y_t, for a cross section of firms at time t, where the inputs into the performance generation process consist of a single endogenous resource, R_t and one exogenous control variable, Z_t. Without loss of generality, we assume that input prices, δ, for resource R_t are known and are equal to 1. Managers pick the level of resources R_t utilized in each time period in order to maximize performance:

$$\max_{Rt} \beta_0 R_t^{\beta_1} Z_t^{\beta_2} - \delta R_t \quad s.t. \quad R_t \le M. \quad (8)$$

Given the managers maximization problem, we can derive a model for resource selection by following the steps outlined above. This involves first writing down the Lagrangian that corresponds to the constrained optimization problems:

$$L = \beta_0 R_t^{\beta_1} Z_t^{\beta_2} - \delta R_t + \lambda(M - R_t). \quad (9)$$

We can then differentiate with respect to the resource and solve for the performance-maximizing level of R_t^*.

$$\frac{\partial L}{\partial R_t} = \beta_0 \beta_1 R_t^{\beta_1 - 1} Z_t^{\beta_2} - \lambda - \delta = 0. \quad (10)$$

We can simplify the expression in equation (10) by taking the natural log of both sides and then solve for the performance maximizing value of the resource:

$$\ln R_t^* = (\ln(\lambda+\delta) - \ln\beta_0 - \ln\beta_1 - \beta_2 \ln Z_t)(\beta_1 - 1)^{-1}. \quad (11)$$

The structural model of optimal resource selection can be transformed into a statistical model by introducing error, ζ, into the resource picking equation:

$$\ln R_t^* = (\ln(\lambda+\delta) - \ln\beta_0 - \ln\beta_1 - \beta_2 \ln Z_t + \zeta_t)(\beta_1 - 1)^{-1}, \quad \zeta_t \sim N(0, \sigma_\zeta^2). \quad (12)$$

This allows us to accommodate deviations from optimality in observed resource utilization and provides us with a formal mechanism to test the hypothesis of endogenous resource selection (e.g luck/other exogenous factors vs. skill). Coupled with the performance model presented in equation (13), Bayesian estimation can proceed as described above:

$$\ln y_t = \ln\beta_0 + \beta_1 \ln R_t^* + \beta_2 \ln Z_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2). \quad (13)$$

By fixing parameter values and simulating data according to our model, we are able to contrast the inference derived from the proposed model of endogenous resource selection with standard estimation approaches used to test the RBV.

Simulation results

Estimates from a cross sectional structural model. In Tables 1–4 and Figure 1 we present the results of a variety of simulation studies that were conducted to ascertain the properties of our proposed structural model relative standard methods of estimation. For each of these studies, data was generated according the process under consideration (e.g., cross-sectional, panel, serial

Table 1 Simulation results for endogenously determined resources

Parameter	*True Value*	*OLS*	*IV*	*Structural Model*
β_0	1.00	1.17 (0.02)	1.16 (0.03)	**0.98** (0.04)
β_1	0.15	−0.16 (0.03)	−0.14 (0.04)	**0.17** (0.07)
β_2	0.55	0.72 (0.03)	0.64 (0.03)	**0.52** (0.04)
λ	0.25	–	–	**0.28** (0.11)
σ_ε^2	0.25	–	–	**0.24** (0.02)
σ_ζ^2	0.25	–	–	**0.24** (0.04)
$\sigma_\varepsilon\sigma_\zeta$	0.10	–	–	**0.09** (0.01)

Notes: Estimated parameters are reported along with standard errors in parentheses.
β_0 = Constant term;
β_1 = Coefficient on the measure of Resources;
β_2 = Coefficient on the control variable;
λ = Shadow price of the resource;
σ_ε^2 = Classical error term for the performance equation;
σ_ζ^2 = Structural error term for the optimization process; and
$\sigma_\varepsilon\sigma_\zeta$ = Covariance between the performance and optimization equation

correlation, etc.) using the "true" parameter values listed in Table 1. For the results presented we generated 100 observations ($T = 100$) from the process described in the model section and then estimate the effects using various techniques that have been used in prior literature and compare it to our proposed estimation approach.[6] As the variable Z is exogenous to the system of study, it is generated from a uniform distribution. Although not presented, this process was repeated for a variety of sample sizes and values of the "true" parameters in order to establish robustness. The general results discussed below were consistent across these various design permutations (see Figure 1 for an example).

Table 1 presents the results of our first simulation study, including the true values of parameters specified for the simulation procedure. Summary statistics for the parameter estimates of our proposed models were computed from the saved draws of the MCMC sampler. Specifically, we ran each MCMC chain for 50K iterations. The first 25K iterations were discarded as burn-in and statistics were computed on the remaining 25K. Estimated parameters whose 95% credible interval contains the corresponding true value are shown in bold face and posterior standard deviations (standard errors) appear in parenthesis. For comparative purposes we contrast the results of our model to the two methods most commonly used to empirically test the RBV, ordinary least squares (OLS) and the method of Instrumental Variables (IV). For the latter, we created an exogenous variable that was correlated with R_t and uncorrelated with ζ_t.[7]

As shown in Table 1, the proposed structural model of endogenous resource picking is able to recover all parameters from the simulation experiment, including the true causal impact of resources on performance as captured by β_1. Not surprisingly, the OLS estimator is unable to recover the causal impact of either R or Z on performance. It overstates the impact of Z and understates that of R. In fact, the estimate of β_1 is negative for OLS, whereas the true value is positive. Interestingly, the results produced from the IV estimator are almost identical to those of the OLS estimator. This result is consistent with the lack of benefits of using standard IV techniques when there is heterogeneity in treatment effects (e.g., Heckman and Robb, 1985; Heckman and Vytlacil, 2005; among others). Taken collectively, this analysis suggests that if managers pick resources with an expectation of how they will influence firm performance and thus introduce heterogeneity in effects, estimation methods that do not explicitly account for this particular form of endogeneity may lead to mis-estimation of the relationship between resources and performance and can actually give rise to sign reversals. In order to generalize the findings from the simulation study presented in Table 1, we conducted the following robustness checks:

First, Table 1 shows results from a model estimated from one realization of data generated from the process in question, governed by a single set of model parameters. To strengthen the analysis, we repeated the simulation and estimation procedure under a variety of circumstances including use of instruments of differing strength (i.e., weak vs. strong correlation with R) and various ranges of associations for β parameters. The general result was consistent across these various permutations.

Figure 1 presents estimated parameter values of β_1 (left panel) and β_2 (right panel) with corresponding 95% confidence intervals for both the OLS (squares) and structural model (circles). In this series of simulations we varied the value of β_1 from 0.15 to 0.75 using an interval length of 0.05. β_2 was computed as a function of β_1 to ensure that adherence to the constraint that the sum of the two parameters be less than 1. In all scenarios the structural model of endogenous resource selection is able to recover the input values of β_1 and β_2 whereas the OLS approach fails to do so. As such, we can conclude that the results of this simulation study seem invariant to the choice of parameter values. Although OLS

[6]We decided to choose the smallest sample in order to illustrate that Bayesian approaches are not plagued by finite sample bias. Small samples can also produce fairly accurate results provided the number of simulations is increased.

[7]It is important to note that we present the best case scenario where a high quality instrument can be identified. In most empirical settings it would be next to impossible to identify quality instruments. It is well established in the literature that the quality of the instrument is critical in determining the value from IV analysis.

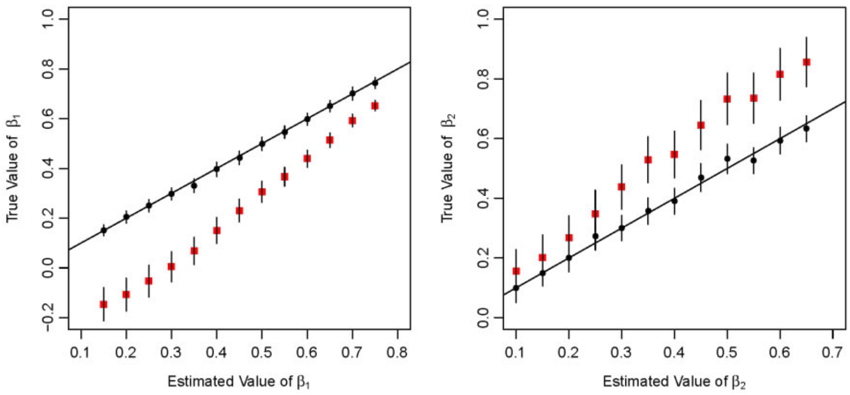

Figure 1 Estimated coefficients of the structural vs. OLS model for varying simulated parameter values
Note: The continuous line corresponds to the true parameters. OLS estimates are denoted by the squares (standard errors by vertical threads). Estimates from the structural model are represented by circles (standard errors by vertical threads).

Table 2 Simulation results when resources are exogenously determined

Parameter	*True Value*	*OLS*	*Structural Model*
β_0	1.25	**1.18** (0.03)	**1.27** (0.03)
β_1	0.25	**0.25** (0.01)	**0.26** (0.01)
β_2	0.60	**0.61** (0.01)	**0.58** (0.01)
λ	–	–	0.01 (0.01)
σ_ε^2	0.05	–	**0.05** (0.00)
σ_ζ^2	–	–	2.54 (0.20)
$\sigma_\varepsilon\sigma_\zeta$	–	–	0.03 (0.03)

Note: Estimated parameters are reported along with standard errors in parentheses, and parameters read as in Table 1.

consistently under-estimates the true value of the endogenous parameter in Figure 1, the direction of mis-estimation does not generalize to all scenarios and is likely specific to these simulation settings. Positive over estimation is also a plausible outcome.

Second, an implicit assumption in the earlier analysis is that managers optimize on their choice of resources. However *ex ante* it is impossible for the analyst to observe this process. To mitigate this concern we investigate the effect of using our proposed structural model on data where managers do not engage in optimal selection of resources or where performance is causally ambiguous (i.e., resources are assumed exogenous to the system of study). Under any of these scenarios, the true data generating mechanism corresponds to consistent estimation using OLS.

The results from the analysis are presented in Table 2. As expected, we find that if resources are exogenous to the system, OLS is able to recover the true effect on resources on performance.[8] Interestingly, we also find that our proposed model of optimal resource picking also recovers the true parameters. As we can see from the results, the direction and statistical significance of the coefficient is preserved in the structural model. The equivalence of the results from OLS and the structural model suggests that even if managers do not optimize and there is no endogeneity, there is no loss associated with using the structural model. Furthermore, the structural model offers the added attraction to the analysis in terms of examining the size of the error variance from the optimization equation σ_ζ^2. As σ_ζ^2 increases, the influence of the resource picking equation on the estimation of β_1 decreases. In the limit, all information about β_1 is contained in the performance equation and estimated parameters from the joint model will be identical to those from the simple model. This can be seen by comparing Tables 1 and 2, the error variance is estimated to be large in the model without endogeneity, suggesting that resources are exogenously determined and that any resulting superior performance can be attributed to alternative explanations like luck, causal ambiguity, socio-political factors or managerial inattention among others. Thus, a comparison of models with and without endogeneity enables researchers to determine whether performance is driven through optimal selection of resources or by some other factor, thus justifying the assumption of exogeneity.

Finally, we need to show that these results do not depend on the choice of a specific functional form. To do so, we specify an alternative functional form that can accommodate additive (as opposed to multiplicative) effects of endogenous resources and control variables. This additive performance equation can be expressed as follows:

$$y_t = \beta_0 + \beta_1 R_t^{\alpha} + \beta_2 Z_t + \varepsilon_t. \tag{14}$$

Where the collection of parameters and variables are used as defined above with the exception of $\alpha \in (0, 1)$,

[8]However, note that the resource to performance linkage though statistically significant overstates the support for the RBV as noted by Arend and Levesque (2010).

Table 3 Simulation results for additive functional form

Parameter	*True Value*	*OLS*	*Structural Model*
β_0	1.25	**1.35** (0.09)	**1.21** (0.07)
β_1	0.25	−10.95 (5.7)	**0.27** (0.07)
β_2	0.60	**0.66** (0.06)	**0.66** (0.06)
λ	0.25	–	**0.25** (0.16)
σ_ε^2	0.25	–	**0.23** (0.02)
σ_ζ^2	0.25	–	**0.24** (0.02)
$\sigma_\varepsilon\sigma_\zeta$	−0.05	–	**−0.03** (0.02)

Note: Estimated parameters are reported along with standard errors in parentheses and parameters read as in Table 1.

which is a fixed (i.e., pre-specified) constant that governs diminishing returns to scale in the efficacy of R_t. If the condition of diminishing returns to scale is met, then we can follow the procedure defined above and derive a model for the selection of an optimal R_t^*. Expressed in logarithmic terms, it can be shown that the optimal value of the endogenous resource given this form of the performance function is:

$$\ln R_t^* = \frac{1}{\alpha - 1}\left(\ln\left(\delta(1+\lambda)\right) - \ln\left(\alpha\beta_1\right)\right) + \zeta_t. \quad (15)$$

Equations (14) and (15) can be used to derive the likelihood for the system and a corresponding Bayesian estimation procedure. Table 3 presents the results of a simulation study conducted using this specification where α was set to 0.5 (i.e., we take the square root of R). Like the former simulation experiments, Table 3 shows that joint modeling of both the performance and resource-picking processes allows us to recover the causal parameters entered in the system, whereas the results generated using OLS yield an incorrectly signed coefficient for the effect of the resource on firm performance, namely,. β_1.

Dynamic effects. The models tested thus far have been specified to accommodate cross-sectional or time series effects where we assume independence across units of time. However, the availability of longitudinal data can allow us to examine more interesting phenomena. In our structural model as specified thus far, we implicitly assume that the costs and benefits from acquisition of a resource accrue in the same time period. This, of course, is a fairly restrictive assumption and may not be justifiable for all types of resources (e.g., Coff, 2010). For example, acquisition of resources such as human capital may result in costs and benefits that accrue in different time periods. Under these circumstances it is possible that given a cross section of the data at time t, the estimated contribution from the resource is underestimated at time t. Since the benefits start accruing only from period $t+k$ we can expect the contribution from the resource to be greater in future periods. This suggests the need for an approach that helps us distinguish between the contemporaneous effect of the resource on performance and its long-run effect.

These carry-over effects can be incorporated into the model by first rewriting the performance equation as:

$$\ln y_{it} = \ln\beta_0 + \beta_1 \ln R_{it}^* + \beta_2 \ln Z_{it} + \varepsilon_{it}, \quad (16)$$

where the subscript i denotes each of the firms in the panel and the error term ε_{it} captures the performance shocks that are not observed by the researcher. Persistence in the effect of the resource acquired in time t on future periods can be captured by allowing the performance shock ε_{it} to be serially correlated over time:

$$\varepsilon_{it} = \rho\varepsilon_{it-1} + u_{it},$$

where u_{it} is the firm specific error term. The estimable specification for the performance function can be written in a dynamic form as below:

$$\ln y_{it} = (1-\rho)\ln\beta_0 + \rho \ln y_{it-1} + \beta_1 \ln R_{it}^* + \beta_2 Z_{it} + u_{it}, \quad (17)$$

where ρ is the persistence parameter that lies in the unit interval [0, 1) and the Resources (R_{it}) and controls (Z_{it}) are inputs from the optimization process as outlined earlier. The model specified above is a standard autoregressive distributed lag (ADL) model. The advantage of this approach is that we can now estimate the long-run multiplier effect. For instance the coefficient β_1 captures the contemporaneous effect of the resource R_{it} on performance. However, if the effect of R_{it} is not fully captured in the same time period the long run multiplier effect can be computed using the formula $\frac{\beta_1}{1-\rho}$ (Davidson and MacKinnon, 1993).

To test this model, we simulate data from the process described above and estimate it using our proposed dynamic model to determine if our approach recovers the true parameters of the data generating mechanism. To maintain consistency across the models we estimate, we restrict the effect of the lagged independent variables to zero and hence exclude it from the dynamic model. We estimate the model on a dataset where resource costs exceed benefits in the initial periods, but where accumulated benefits ultimately dominate costs, thus justifying acquisition of the resource. Results from our analysis are presented in Table 4.

We also apply traditional pooled OLS and panel data methods to this particular data set. The latter is implemented by fitting a random coefficients model (i.e., a Bayesian hierarchical linear model). In both the OLS and panel approach we find that the resource to performance coefficient is negative, suggesting that the unit

Table 4 Simulation results with serially auto-correlated errors

Parameter	*True Value*	*Pooled Model*	*Panel Model*	*Structural Model*
β_0	1.00	0.74 (0.05)	0.55 (0.06)	**0.98** (0.04)
β_1	0.15	−0.17 (0.04)	−0.17 (0.03)	**0.16** (0.03)
β_2	0.55	0.83 (0.04)	0.84 (0.04)	**0.55** (0.03)
λ	0.25	–	–	**0.26** (0.19)
ρ	0.20	–	0.20 (0.03)	**0.19** (0.03)
σ_ε^2	0.25	0.47 (0.01)	0.45 (0.01)	**0.25** (0.02)
σ_ζ^2	0.25	–	–	**0.25** (0.02)
$\sigma_\varepsilon\sigma_\zeta$	0.10	–	–	**0.09** (0.01)

Note: Estimated parameters are reported along with standard errors in parentheses. Parameters read as in Table 1 except for ρ which captures the degree of autocorrelation of performance across time.

addition of the resource decreases performance. The result is not surprising given that costs exceed benefits for a single period taken in isolation. As was the case with IV estimation, controlling for unobserved heterogeneity through fixed or random effects is not helpful in solving issues posed by structural endogeneity. From this analysis, we conclude that simple methods that ignore the long-term impact of the resource can yield estimates of the relationship between resource and performance that are negative even when the long-term net benefit is positive.

Results from the structural estimation approach, however, appropriately capture the dynamics of the model. Specifically, we are able to recover the true estimate of the persistence parameter thus allowing us to compute long-term benefit of the resource. Lower values of the persistence parameter ρ indicate that the cumulative effect of the resource is quickly absorbed in performance and vice versa. We also find that the effect of the resource on performance is also positive, thus implying that the structural model factors in the private information (through the error variance of the optimization equation) in the estimation process and hence recovers the true underlying parameters of the model.

Discussion

We introduced this paper by suggesting that while the RBV is indeed a dominant theoretical framework in the field of strategy, empirical evidence in support of the theory is mixed. This lack of consensus can be attributed to the incapacity of empirical designs to capture endogeneity arising out of resource picking decisions. Extant literature provides us with formal models about the resource picking behavior of firms indicating the presence of endogenous choices (e.g., Barney, 1986; Makadok, 2001). While the concept of endogeneity and correcting for it is not new to the discipline of strategic management (e.g., see Hamilton and Nickerson, 2003), an aspect that is often ignored is determining the source of endogeneity, which can potentially be embedded in theory. Recent literature in program evaluation shows that using popular endogeneity correction techniques such as IV based approaches (e.g., 2 stage least squares), panel data models, dynamic panel models (GMM based approaches) and treatment effect models (e.g., sample selection, matching, etc.) work well only when certain underlying assumptions regarding the source of endogeneity is satisfied. Of critical importance is the assumption that the effect of the endogenous variable is homogenous across firms and also that choices occur as if managers do not use their skills in identifying value enhancing resources.

However, we argue that this situation does not fit well with the RBV framework. First, the RBV is a theory about heterogeneity and implies that certain firms are indeed superior to others. Second, prior literature suggests a significant role for resource picking skills to have an impact on performance (Barney, 1986; Makadok, 2001). Further, from a dynamic perspective, literature suggests that complementarities between existing/future resources may also play a critical role in superior performance or governance choices (e.g., Argyres and Zenger, 2012). Within the dynamic perspective, this has two implications. First, the same resource may be valued differently by firms depending on existing or future expected complementarities with other resources. Second, managers and firms are more than likely to factor in these complementarities when designing their resource configurations. In this backdrop, we offer a potential solution to the "essential endogeneity" problem posed by the resource picking decision and demonstrate through simulations the validity of our approach.

The foundations of the model are built on a simple input-output process, where the inputs (resources) impact the output (performance). The resources are the outcome of an optimization process that the firms (managers) engage in. This optimization process is based up choosing the best possible resource configuration within the constraints imposed by the RBV such that it maximizes firm performance. Through the introduction of the optimization process, we address the key concern of the sorting gain induced by endogeneity into the model.

While our core model is developed for estimation using data from a cross section of firms, the extension to longitudinal data is relatively simple to implement. This is accomplished by connecting individual, firm-level models through a hierarchical structure. The general version of our model presented in equations (1) and (3) is specified for data with a panel structure and a corresponding estimation strategy appears in the appendix of Dotson and Allenby (2010). Further details on Bayesian hierarchical modeling are presented in Gelman and Hill (2007). The extension to panel data can provide a variety of insights. For example, persistence in performance can be directly modeled using longitudinal data, thus addressing the concern regarding the effects of resources on sustainable performance, as shown in Table 4. Furthermore, while the costs incurred in acquiring a resource and the benefits from the resource may accrue in the same period, it is not always true for the entire portfolio of resources. With the help of longitudinal data, researchers can now identify long-run vs. short-run effects of resources on performance, thus allowing for various types of resource effects to be tested.

The data requirement for analysis using our model mirrors that of any normal RBV study, that is, measures of performance which in this case must be positive, observed resource configurations and necessary control variables. The imposition of constraints emerges from the operationalization of the shadow prices, which are estimated from the model. For instance, the constraint on inputs (*M* in equation 8) need not be physically observed. The structure of the model is preserved with the weak assumption that there exists a positive level (*M*) that serves as the limit beyond which resources are not available.

One challenge in implementing our proposed model (relative to standard approaches) is that commercial software packages do not exist to run the models described above. Estimation for our model was conducted using the R project for statistical computing where the MCMC sampler was coded by hand as shown in the appendix.[9] Given an identical data structure to the one we simulate (i.e., one endogenous resource and one control variable), this sampler could be applied to a new dataset. If the data structure differs, the general process for model derivation and estimation would be to:

1. Specify a functional relationship (e.g., a linear relationship in a typical regression model) between resources and performance. The nature of this relationship should be motivated by both theory and the structure of available data.
2. Write down a managerial optimization problem for the endogenous variables. This involves determining what management is trying to accomplish, which variables they control and the relevant set of constraints they face. Given a well articulated optimization problem, compute optimal utilization of the endogenous resource using the method of Lagrange multipliers described above. The resulting equation(s) will form the basis of the resource-picking model.
3. Derive the likelihood for the joint system given assumptions about the error distribution for both the performance and resource-picking equations.
4. Specify priors for all model parameters and sample from the resulting posterior distribution using MCMC methods.

Although this process is more challenging to implement than simply applying a standard model, the resulting benefits of improved inference and added insights are worth the effort. The fact that the model is derived from theory allows the researcher to test a variety of qualitatively distinct model specifications while accounting for the unique features of a particular dataset.

While the model prescribed in this paper can be applied to a wide variety of resources, it still has some constraints. For instance, some resources may evolve over time with performance, such as experience or reputation. Given the structural nature of our model and to be consistent with the evolving nature of these resources, there is a need for specifying a dynamic evolution path for such resources. While the model does not currently accommodate for these feedback relationships, it opens up an avenue for future research in this direction. Notably, the Bayesian approach outlined in this paper can be modified to allow the researcher to specify multi-level models, where a second level could specify a dynamic process for the evolution of specific capabilities. Also, ways to compute inter-resource complementarity needs further investigation so that comparison of effects between managerial picking and heterogeneous resource complementarity can be tested (Adegbesan, 2009). We leave these promising ideas for future work.

Finally, our model proves to be superior to what can be obtained by direct OLS models and IV models based on simulated data. Given the analytical results provided by Heckman *et al*. (2006); issues of erroneous inference and endogeneity characterizing resource picking as always emanating from a localized context (resource is not exogenous) need particular methodological treatments. The possibility of testing further counterfactual propositions is a natural prolongation of the model proposed here, addressing thorny concerns of causation (Durand and Vaara, 2009; Tang and Liou, 2010; Makadok, 2011).

Overall, taking stock of theoretical debates around RBV, our study joins the efforts of others especially with the recent call to introduce Bayesian methods into the field of management (Hansen *et al*., 2004; Hahn and Doh, 2006 and Kruschke *et al*., 2012) to model the causal relationship between resources and superior performance. By so doing, we aim to participate in the edification of strategy as a scientific discipline established around a limited number of coherent propositions, united into what the sociology of science calls a paradigm, which will be of interest to other disciplines beyond management. Already, in association with other techniques (stochastic frontier approaches, other

[9]We provide the R-code to run our simulation models in Appendix B.

Bayesian models), the structural model of resource-picking and performance proposed here advances our movement toward a more firmly established paradigm of strategic management, both theoretically and empirically.

References

Adegbesan, J. A., 2009, "On the origins of competitive advantage: Strategic factor markets and heterogenous resource complementarity". *Academy of Management Review*, **34**: 463–475.

Argyres, N. and T. Zenger, 2012, "Capabilities, transaction costs and firm boundaries". *Organization Science*, **23**: 1643–1667.

Arend, R. J., 2006, "Tests of the resource-based view: Do the empirics have any clothes?" *Strategic Organization*, **4**: 409–422.

Arend, R. J. and M. Levesque, 2010, "Is the resource-based view a practical organizational theory? " *Organization Science*, **21**: 913–930.

Arrelano, M. and S. R. Bond, 1991, "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations". *Review of Economic Studies*, **58**: 277–297.

Barney, J. B., 1986, "Strategic factor markets: Expectations, luck and business strategy". *Management Science*, **32**: 1231–1241.

Barney, J. B., 1991, "Firm Resources and Sustained Competitive Advantage". *Journal of Management*, **17**: 99–120.

Barney, J. B. and A. M. Arikan, 2001, "The resource-based view: origins and implications". In M. A. Hitt, R. E. Freeman and J. S. Harrison (eds.) *The Blackwell handbook of strategic management*. Oxford: Blackwell, 124–188.

Bascle, G., 2008, "Controlling for endogeneity with instrumental variables in strategic management research". *Strategic Organization*, **6**: 285–327.

Bierly, P. and A. Chakrabarti, 1996, "Generic knowledge strategies in the US pharmaceutical industry". *Strategic Management Journal*, **17**: 123–136.

Blundell, R and S. Bond, 1998, "Initial conditions and moment restrictions in dynamic panel data models". *Journal of Econometrics*, **87**: 115–143.

Certo, T. and M. Semadeni, 2006, "Strategy research and panel data: Evidence and implications". *Journal of Management*, **32**: 449–471.

Chang, X. and S. Dasgupta, 2009, "Target behavior and financing: How conclusive is the evidence?" *Journal of Finance*, **64**: 1767–1796.

Chatterjee, S. and B. Wernerfelt, 1991, "The link between resources and type of diversification: Theory and evidence". *Strategic Management Journal*, **12**: 33–48.

Chib, S. and E. Greenberg, 1995, "Understanding the Metropolis-Hastings algorithm". *American Statistician*, **49**: 327–335.

Coff, R., 2010, "The co-evolution of rent appropriation and capability development". *Strategic Management Journal*, **31**: 711–733.

Collis, D. J., 1994, "How valuable are organizational capabilities". *Strategic Management Journal*, **15**: 49–68.

Conley, T., C. Hansen and P. E. Rossi, 2012, "Plausibly exogenous". *The Review of Economics and Statistics*, **94**: 260–272.

Crook, T. R., D. J. Ketchen Jr., J. G. Combs and S. Y. Todd, 2008, "Strategic resources and performance: A meta-analysis". *Strategic Management Journal*, **29**: 1141–1154.

Davidson, R. and J. G. MacKinnon, 1993, *Estimation and inference in econometrics*. New York: Oxford University Press.

Denrell, J., 2004, "Random walks and sustained competitive advantage". *Management Science*, **50**: 922–934.

Denrell, J. and C. Fang, 2010, "Predicting the next big thing: Success as a signal of poor judgment". *Management Science*, **56**: 1653–1667.

Dotson, J. P. and G. M. Allenby, 2010, "Investigating the strategic influence of customer and employee satisfaction on firm financial performance". *Marketing Science*, **29**: 895–908.

Durand, R., 2002, "Competitive advantages exist: A critique of Powell". *Strategic Management Journal*, **23**: 867–872.

Durand, R. and E. Vaara, 2009, "Causation, counterfactuals and competitive advantage". *Strategic Management Journal*, **30**: 1245–1264.

Erickson, T and T. M. Whited, 2002, "Two-step GMM estimation of the errors-in-variables model using higher order moments". *Econometric Theory*, **18**: 776–799.

Fornell, C. and D. Larcker, 1981, "Structural equation models with unobservable variables and measurement error". *Journal of Marketing Research*, **18**: 39–50.

Foss, N. J. and N. Hallberg, 2010, "Symmetry and theoretical isolation in the resource-based view". Working Paper, Copenhagen Business School

Gelman, A. and J. Hill, 2007, *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.

Godfrey, P. C. and C. W. L. Hill, 1995, "The problem of unobservables in strategic management research". *Strategic Management Journal*, **16**: 519–533.

Hahn, E. D. and J. P. Doh, 2006, "Using Bayesian methods in strategy research: An extension of Hansen *et al*.". *Strategic Management Journal*, **27**: 783–798.

Hamilton, B. H and J. A. Nickerson, 2003, "Correcting for endogeneity in strategic management research". *Strategic Organization*, **1**: 51–78.

Hansen, M. H., L. T. Perry and S. C. Reese, 2004, "A Bayesian operationalization of the resource-based view". *Strategic Management Journal*, **25**: 1279–1295.

Heckman, J. J., 1979, "Sample selection bias as a specification error". *Econometrica*, **47**: 153–161.

Heckman, J. J and R. Robb, 1985, "Alternative methods for estimating the impact of interventions". In J. Heckman and B. Singer (eds.) *Logitudinal analysis of labor market data*. New York: Cambridge University Press, 156–245.

Heckman, J. J and E. Vytlacil, 1998, "Instrumental variable methods for the correlated random coefficient model: Estimating the average rate of return to schooling when the return is correlated with schooling". *Journal of Human Resources*, **33**: 974–1002.

Heckman, J. J and E. Vytlacil, 2005, "Structural equations, treatment effects and econometric policy evaluation". *Econometrica*, **73**: 669–738.

Heckman, J. J., S. Urzua and E. Vytlacil, 2006, "Understanding instrumental variables in models with essential heterogeneity". *The Review of Economics and Statistics*, **88**: 389–432.

Henderson, A. D., M. E. Raynor and M. Ahmed, 2012, "How long must a firm be great to rule out chance? Benchmarking superior performance without being fooled by randomness". *Strategic Management Journal*, **33**: 387–406.

Henderson, R. and I. Cockburn, 1994, "Measuring competence? Exploring firm effect in pharmaceutical research". *Strategic Management Journal*, **15**: 63–84.

Imbens, G. W and J. D. Angrist, 1994, "Identification and estimation of local average treatment effects". *Econometrica*, **62**: 467–475.

Kotha, S., 1995, "Mass customization: Implementing the emerging paradigm for competitive advantage". *Strategic Management Journal*, **16**: 21–42.

Kruschke, J. K., H. Aguinis and H. Joo, 2012, "The time has come: Bayesian methods for data analysis in the organizational sciences". *Organizational Research Methods*, **15**: 722–752.

Makadok, R., 2001, "Toward a synthesis of the resource-based and dynamic-capability views of rent creation". *Strategic Management Journal*, **22**: 387–402.

Makadok, R., 2011, "The four theories of profit and their joint effects". *Journal of Management*, **37**: 1316–1334.

Miller, D. and J. Shamsie, 1996, "The resource-based view of the firm in two environments: the Hollywood film studios from 1936 to 1965". *Academy of Management Journal*, **39**: 519–543.

Mowery, D. C., J. E. Oxley and B. S. Silverman, 1996, "Strategic alliances and inter-firm knowledge transfer". *Strategic Management Journal*, **17**: 77–92.

Newbert, S. L., 2007, "Empirical research on the resource-based view of the firm: An assessment and suggestions for future research". *Strategic Management Journal*, **28**: 121–146.

Newbert, S. L., 2008, "Value, rareness, competitive advantage and performance: A conceptual-level empirical investigation of the resource-based view of the firm". *Strategic Management Journal*, **29**: 745–768.

Otter, T., T. J. Gilbride and G. M. Allenby, 2011, "Testing models of strategic behavior characterized by conditional likelihoods". *Marketing Science*, **30**: 686–702.

Park, S. and S. Gupta, 2012, "Handling endogenous regressors by joint estimation using copulas". *Marketing Science*, **31**: 567–586.

Penrose, E. T., 1959, *The theory of the growth of the firm*. New York: John Wiley.

Powell, T. C., 2001, "Competitive advantage: Logical and philosophical considerations". *Strategic Management Journal*, **22**: 875–888.

Prabhala, N., 1997, "Conditional methods In event-studies and an equilibrium justification for using standard event-study methods". *Review of Financial Studies*, **10**: 1–38.

Priem R. L. and J. E. Butler, 2001, "Is the resource-based 'view' a useful perspective for strategic management research?" *Academy of Management Review*, **26**: 22–40.

Rigobon, R., 2003, "Identification through heteroskedasticity". *The Review of Economics and Statistics*, **85**: 777–792.

Rosenbaum, P. R. and D. B. Rubin, 1983, "The central role of the propensity score in observational studies for causal effects". *Biometrika*, **70**: 41–55.

Rossi, P. E., G. M. Allenby and R. McCulloch, 2005, *Bayesian statistics and marketing*. New York: John Wiley and Sons.

Semadeni, M., M. C. Withers and S. T. Certo, 2014, "The perils of endogeneity and instrumental variables in strategy research: Understanding through simulations". *Strategic Management Journal (Forthcoming)*.

Shaver J. M., 1998, "Accounting for endogeneity when assessing strategy performance: does entry mode choice affect FDI survival?" *Management Science*, **44**: 571–585.

Shaver, M. J., 2007, "Interpreting empirical results in strategy and management research". *Research Methodology in Strategy and Management*, **4**: 273–293.

Swait, J. and J. Louviere, 1993, "The role of the scale parameter in the estimation and comparison of multinomial logit models". *Journal of Marketing Research*, **30**: 305–314.

Szulanski, G., 1996, "Exploring internal stickiness: Impediments to the transfer of best practice within the firm". *Strategic Management Journal*, **17**: 27–43.

Szulanski, G., 1999, "The process of knowledge transfer: A diachronic analysis of stickiness". *Organizational Behavior and Human Decision Processes*, **82**: 9–27.

Tang, Y.-C. and F.-M. Liou, 2010, "Does firm performance reveal its own causes? The role of Bayesian inference". *Strategic Management Journal*, **31**: 39–57.

Whited, T. M and M. R. Roberts, 2012, "Endogeneity in empirical corporate finance". In G. M. Constantinides, M. Harris and R. M. Stulz (eds.) *Handbook of the economics of finance*. Amsterdam: Elsevier, pp. 493–572.

Wiggins, R. R. and T. W. Ruefli, 2002, "Sustained competitive advantage: Temporal dynamics and the incidence and persistence of superior economic performance". *Organization Science*, **13**: 82–105.

Yang, S., Y. Chen and G. M. Allenby, 2003, "Bayesian analysis of simultaneous demand and supply". *Quantitative Marketing and Economics*, **1**: 251–275.

Zollo, M. and S. G. Winter, 2002, "Deliberate learning and the evolution of dynamic capabilities". *Organization Science*, **13**: 339–351.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Online Apendix. R-CODE to estimate our example structural model

Appendix A

Estimation algorithm

Our proposed models are estimated using Bayesian statistical methods. Bayesian inference is accomplished by generating many realizations (i.e., "draws") from the posterior distribution of the model parameters, where the posterior distribution is proportional to the prior distribution multiplied by the likelihood of the data:

$$\pi(\theta \mid data) \propto \pi(\theta)\ell(data \mid \theta)$$

Through repeated sampling we able to generate an empirical approximation to the posterior distribution and can use the resulting collection of draws to compute a variety of quantities of interest like point estimates of parameters and confidence intervals.

Modern computational techniques for Bayesian inference focus on finding ways to sample from probability distributions (i.e., the posterior distribution) which may be of unknown parametric form. To this end, a variety of techniques have been developed. In our application, we employ two of these techniques: Gibbs sampling and the Metropolis-Hastings (M-H) algorithm. Gibbs sampling is employed when the prior and likelihood are of known parametric form and are conjugate to one another. This occurs when the product of the prior and likelihood is also of know parametric form. For example, if the likelihood and prior are both normally distributed then the posterior distribution will also be normal. The M-H algorithm is used when the prior and likelihood are not conjugate and, as a result, the form of the posterior distribution is unknown. It provides a series of steps that can be used to sample from the posterior in these cases. We refer the interested reader to the excellent description of the algorithm presented in Chib and Greenberg (1995).

Bayesian estimation of the simplistic model of endogenous resource selection presented in equation (8) proceeds by recursively generating draws from the full conditional distributions of all model parameters. The likelihood for this model can be written as:

$$\ell(\text{data} \mid \text{else}) = \prod_t \pi\left(\hat{\varepsilon}_t, \hat{\zeta}_t\right) \left| J_{t} \right|_{\hat{\zeta} \to \ln(R_t)}$$

where $\pi(\cdot)$ denotes evaluation of the multivariate Normal density, $\left(\hat{\varepsilon}_t, \hat{\zeta}_t\right) \sim N(\underset{\sim}{0}, \Sigma)$ and $\Sigma = \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_\varepsilon \sigma_\zeta \\ \sigma_\varepsilon \sigma_\zeta & \sigma_\zeta^2 \end{bmatrix}$. Additional quantities of interest include:

$$\hat{\varepsilon}_t = \ln(y_t) - \ln(\beta_0) - \beta_1 \ln(R_t) - \beta_2 \ln(Z_t),$$

$$\hat{\zeta}_t = (\beta_1 - 1)\ln(R_t) - \left(\ln(\lambda + 1) - \ln(\beta_0) - \ln(\beta_1) - \beta_2 \ln(Z_t)\right),$$

$$\left| J_{t} \right|_{\hat{\zeta} \to \ln(R_t)} = |\beta_1 - 1|$$

Step 1: draw $[\beta \mid else]$
β is drawn via the Metropolis-Hastings algorithm (see Rossi *et al.*, 2005, p. 88) where the likelihood contribution is defined above and the prior on β is specified as $\exp\{\beta\} \sim N(0, 100I_3)$, where I_3 denotes the identity matrix of dimension 3.

Step 2: draw $[\lambda \mid else]$
λ is also drawn using the Metropolis-Hastings algorithm where the likelihood contribution is identical to step 1 and prior specified as $\lambda \sim N(1, 100)$.

Step 3: draw $[\Sigma \mid else]$
Conditional upon realizations β and λ, Σ is drawn from an Inverted Wishart distribution:

$$\Sigma \sim IW(v_0 + T, V_0 + S)$$

where T is the number of observations, $v_0 = 5$ are the prior degrees of freedom, V_0 is specified as $V_0 = .01I_2$ and S is computed as $S = \underset{2\times T}{\begin{bmatrix} \hat{\varepsilon} \\ \hat{\zeta} \end{bmatrix}} \underset{T\times 2}{\begin{bmatrix} \hat{\varepsilon} & \hat{\zeta} \end{bmatrix}}$.