Evolutionary Trajectories in Petroleum Firm R&D

Constance E. Helfat
Department of Management, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Tacit knowledge and cumulative learning underlie an evolutionary theory of business firm development and strategy. As one test case of the theory, this study examines firms' applied research and development activities. Evolutionary theory suggests that firms within an industry will tend both to persist and to differ in the amount of effort they devote to various R&D applications. A test of the hypothesis of persistent differences in R&D, using uniquely detailed data from the petroleum industry, provides support for evolutionary theory.
(*Innovation; Evolution; Research and Development*)

1. Introduction

Based on an evolutionary theory of business firm development as formulated by Nelson and Winter (1982), this study tests propositions about organizational learning and know-how as they apply to the research and development (R&D) activities of business firms. Applied R&D in particular tends to involve cumulative learning and tacit knowledge, which are characteristics that lie at the heart of an evolutionary theory of business firm development and strategy (Winter 1987). For this reason, an examination of firms' applied R&D activities provides an excellent test case of the applicability of evolutionary theory to the internal functioning of the firm. If data on applied R&D do not support the theory, it is unlikely to apply to other of firms' activities.

An evolutionary theory of business firm development and economic change incorporates both the economic "natural selection" of firms and "organizational genetics" (Nelson and Winter 1982, p. 9). In an analogy to natural selection in biology, economic natural selection refers to the role of the external market environment in determining the survival and growth of firms. Organizational genetics deals with the transmission of organizational traits through time. Nelson and Winter's evolutionary theory is Lamarckian; firms inherit acquired characteristics, and also can adapt and change. Nevertheless, organizations are "typically much better at the tasks of self-maintenance in a constant environment than they are at major change, and much better at changing in the direction of "more of the same" than they are at any other kind of change" (Nelson and Winter 1982, pp. 9–10).

This study focuses on the organizational genetics side of a Lamarckian evolutionary theory of business firm development and strategy, and investigates two complementary propositions of such an evolutionary theory: (1) an individual firm tends to pursue activities similar to those undertaken in the past, and (2) firms operating in a particular market (or set of markets) differ from one another in the activities they undertake. With regard to corporate applied R&D, these propositions suggest that firms *within* an industry will *persistently differ* in the amount of effort, and thus expenditures, devoted to various R&D applications.

Few empirical studies have focused simultaneously on the persistence of firms' R&D activities and on differences between firms in their R&D endeavors, particularly within industries. This study does both. By disaggregating firms' R&D expenditures by type of business application, this study can more finely analyze persistent differences *between* firms in each R&D application than is possible with only *total* firm R&D data utilized in most studies. The study also can provide evidence regarding persistent differences *within* firms in the effort

0025-1909/94/4012/1720$01.25

devoted to various R&D applications. An empirical analysis of uniquely detailed data on petroleum firms' R&D expenditures supports the twin propositions from evolutionary theory of persistence and differences in firms' activities.

This study builds on a prior exploratory analysis, which provided descriptive evidence of relatively stable differences between petroleum firms in their R&D applications and expenditures (Helfat 1994). The prior study focused in detail on the characteristics of highly applied R&D, such as that in the petroleum industry, which tend to tie the R&D to the firm in which it occurs—making the R&D firm-specific. In contrast, this study focuses more broadly on evolutionary theory and provides statistical tests of hypotheses regarding persistent differences in R&D.

Recently, some scholars (e.g., Meyer 1990 and Singh 1990) have called for greater study of intraorganizational evolution. Although organization behavior scholars have empirically analyzed the evolution of populations of firms, researchers have devoted much less attention to empirical analysis of the evolution of internal firm capabilities. Mainstream industrial organization economists also have focused little attention on the evolution of competencies, either in general or regarding R&D. This study addresses itself directly to the evolution of firms' R&D competencies, as reflected in the pattern through time of firms' underlying R&D expenditures. The study has implications for strategic management, since an understanding of the extent to which firms' capabilities both persist through time and differ across firms may increase our understanding of firms' long-term competitive abilities.

The paper proceeds as follows. Section 2 discusses concepts from evolutionary theory relevant to the study of R&D, and presents testable hypotheses which taken together point to persistent intraindustry differences in firms' R&D activities. Section 3 compares evolutionary theory with alternative explanations from neoclassical economics of persistence within and differences between firms' R&D. Section 4 briefly discusses similarities between evolutionary theory and the resource-based view in strategic management. Section 5 then describes the data used to test the hypotheses from §2, §6 contains the empirical analysis, and §7 summarizes the findings and their implications.

2. Organizational Learning, Evolutionary Theory, and R&D

The preponderance of R&D by business firms consists of applied and developmental research (directed primarily towards the commercialization of technologies and scientific discoveries) as opposed to basic (i.e., scientific) research. In general, the ability to develop and design new products and processes involves the ability to learn (Dosi et al. 1992). Many authors have noted the cumulative, often tacit and firm-specific nature of organizational learning (see, e.g., Cyert and March 1992, Dosi 1988, Dosi et al. 1992, Cohen and Levinthal 1990, Nelson 1990, and Nelson and Winter 1982). The discussion in this section deals with each of these characteristics of organizational learning in turn, and presents hypotheses regarding the pattern of firms' R&D endeavors within an industry. In order to focus on similarities and differences between firms rather than between external market environments, the following discussion concentrates on corporate applied R&D activities directed toward similar ends in similar market environments. The empirical example that will be examined shortly is applied R&D on energy-related technologies directed toward fuel extraction, fuel refining, and new fuel sources.

2.1. Persistence in R&D

Perhaps the most obvious characteristic of organizational learning is that it is cumulative. Cohen and Levinthal (1990) note that research on memory development suggests that accumulated prior knowledge increases both the ability to put related, new knowledge into memory, and the ability to recall and use it. This further suggests that for individuals, "learning performance is greatest when the object of learning is related to what is already known" (Cohen and Levinthal 1990, p. 131). Since organizational learning involves the joint contributions of individuals directed toward complex problems (Dosi et al. 1992), an organization's ability to learn depends at least in part on the ability of its individual members to learn. It follows that, like individual learning, organizational learning at least to some extent is cumulative as well.

The cumulative nature of organizational learning has implications for the search for new knowledge. Specifically, firms search "locally," in the neighborhood of current knowledge (Cyert and March 1992, Nelson 1990, Nelson and Winter 1982). In searching for new products and processes, a firm does not examine all possible alternatives, for two reasons. First, bounded rationality (Simon 1978) makes it difficult to uncover all possible options, and to accurately evaluate their future prospects. But more importantly, because possession of an accumulated knowledge base facilitates learning related to that knowledge, a firm often searches for new products and processes in the areas where it has current expertise.

The cumulative nature of learning and local search for new knowledge both imply that firms build on past areas of expertise. Firms tend to concentrate their R&D efforts in areas related to preexisting knowledge bases, and tend to produce new knowledge closely related to the old. In addition, if standard operating procedures and routines for internal allocation of funds are relatively stable through time, then budgetary allocations for R&D will depend strongly on previous budgetary allocations, thus reinforcing any persistence in R&D caused by cumulative learning (Cyert and March 1992, Nelson and Winter 1982). Persistence in R&D is an example of a wider phenomenon termed "path dependence," broadly defined as the tendency to pursue activities similar to those undertaken in the past. (See David (1988) for a detailed discussion of path dependence.) In the absence of large or rapid changes in the external selection environment or in the firm's internal learning environment (such as unanticipated technological bottlenecks or major internal organizational changes), path dependence produces patterns of R&D activity that reflect continuing efforts in the same technological direction(s) as in the past.

For purposes of this analysis, an R&D activity is defined as that directed toward a particular type of business application (e.g., oil and gas recovery R&D or oil refining R&D). Using this definition, H1 is the first hypothesis regarding persistence in R&D to be examined empirically.

H1: *A firm's current and past efforts in each R&D activity are positively correlated.*

Evolutionary theory suggests that any observed path dependence in R&D over time reflects incremental, localized adaptation as firms search for new knowledge primarily in the neighborhood of their current knowledge bases (Nelson and Winter 1982, Winter 1990). Alternatively, path dependence in R&D could result from firm policies that do not change through time, as in many organizational ecology models. Hannan and Freeman (1989) note that sunk costs (such as past accumulated R&D effort) constrain adaptation. In an effort to reconcile the hypothesis of infrequent organization change with that of organizational adaptation, some researchers have suggested that peripheral features of organizations adapt more easily than do core features (Scott 1987, Hannan and Freeman 1989). The fact that R&D laboratories "are involved in the production or management of economic change as their *principal* function" (Nelson and Winter 1982, p. 97), however, suggests that incremental change characterizes R&D regardless of its centrality to the organization.

Ultimately, the degree of adaptation is an empirical question (Hannan and Freeman 1989, Winter 1990). If H1 is correct and current and past R&D activities are positively correlated, the prior discussion suggests this correlation could be due to: a) slowly but nevertheless incrementally changing R&D activity over time, or b) unchanging R&D activity over time. Following evolutionary theory, H2 posits incremental change.

H2: *If current and past efforts in an R&D activity within a firm are positively correlated, this results from a process of incremental adjustment of R&D activity over time.*

2.2. Differences in R&D

Evolutionary theory also emphasizes the often tacit and firm-specific nature of organizational knowledge. Tacit knowledge refers to knowledge that cannot be written down in "blueprint" form, easily transferable between people and organizations (Dosi 1988, Kogut 1988, Nelson and Winter 1982, Polanyi 1967). Tacitness can produce firm-specificity of organizational knowledge: to the extent that the tacit nature of organizational knowledge makes it more difficult to transfer knowledge between firms or to reproduce it by imitation, the knowledge remains specific to the firm that created it. In addition, organizational learning typically requires common codes of communication to facilitate coordination between the individuals involved, and these communication codes tend to be firm-specific (Dosi et al. 1992, Nelson and Winter 1982).

The tacit and firm-specific nature of organizational knowledge and learning impedes the transfer of knowledge between firms. If firms come into existence with different initial knowledge bases, if organizational learning is path-dependent, and if the transfer across firms of organizational knowledge obtained via path-dependent R&D is difficult (or even just imperfect), then firms may not have the same R&D-related knowledge bases. Firms also may differ in the R&D activities they pursue, as they build on their differential knowledge bases in a path-dependent manner. As Nelson (1991) observes, under evolutionary theory, firms have different capabilities, including R&D capabilities, and pursue somewhat different paths. (See also Dosi 1988.) The next hypothesis follows from this logic.

H3: *For a given R&D activity, the level of a firm's effort differs from that of other firms.*

In addition to differences in R&D activities *between* firms, evolutionary theory contains the implication that the levels of effort expended on various R&D activities will *differ within firms*. The logic is as follows. Because for any R&D activity the probability of future success is uncertain, a firm may spread the risk of failure by pursuing more than one R&D activity. When R&D costs are small relative to capital investment in new technologies, a "parallel-path" approach (Nelson 1961) of conducting several R&D activities allows firms to gain greater information about the costs and probabilities of success of various technologies prior to making large investments in risky projects. In employing a parallel-path approach to R&D, a company must decide which types of R&D to pursue and how much effort to devote to each R&D activity. Factors external to the firm may influence firms' choices. For example, market information relevant to the prediction of future prices for the output of R&D or the state of development of basic science may affect firms' estimates of the returns to various R&D activities. These external factors, common to all firms, could cause differences within firms in the amounts of effort devoted to different R&D activities. Such reasoning suggests that all firms would have the same pattern of within firm differences in R&D effort. Factors internal to the firm, however, also affect within firm allocations of R&D effort. The earlier discussion of path dependence suggests that a firm may pursue most strongly those R&D activities in which it has greater accumulated knowledge. If firms' knowledge bases differ and firms conduct R&D that builds on their different knowledge bases, the pattern of within firm differences in R&D also will differ between firms.

H4: *Firms tend to emphasize areas in which they have accumulated knowledge in the past. An individual firm does not necessarily devote the same degree of R&D effort to all R&D activities undertaken by firms in the industry. The pattern of within firm differences in R&D effort also differs between firms.*

Thus far, the discussion has focused primarily on aspects of evolutionary theory having to do with internal firm organization. Hypotheses 1 through 4 in combination point to *persistent differences* in firms' R&D activities within an industry. Feedback from the external selection environment also may affect firms' R&D activities. For example, major changes in the market environment (such as large changes in product or input prices) may temporarily decrease the path dependence of R&D activities as firms respond by altering the composition or amount of their R&D efforts. The eventual impact on the path dependence of different types of R&D activities depends on the interaction of the particular environmental changes with a firm's existing knowledge bases in various R&D activities and with a firm's budgetary routines. A firm might reallocate funds toward those activities thought to have the greatest combined favorable market environment *and* probability of technological success for the firm; the latter partly depends on the knowledge base accumulated from past R&D efforts. Obviously, the interaction of changes in the external environment with internal firm activities is a complex process, and one that is more difficult to examine empirically than the previous hypotheses. Although the empirical analysis in this paper will discuss feedback from the external selection environment, it proves difficult to directly test hypotheses regarding feedback from the market using the petroleum industry R&D data in this study.

3. Neoclassical Economics vs. Evolutionary Theory

In setting forth an evolutionary theory of economic change, Nelson and Winter (1982) had two major aims: to focus attention on economic change, and to argue for "a major reconstruction of the theoretical foundations of. . . orthodox [economic] theory of firm and industry behavior" (Nelson and Winter 1982, p. 4). Hypotheses 1 through 4 proposed here obviously deal with the issue of economic change. The issue arises, however: do these hypotheses conflict with orthodox economic explanations of business firm R&D? Given that so much academic research on industrial R&D has taken place in economics, this question bears at least brief examination.

The general proposition of persistent differences between firms contrasts strongly with the simplest form of neoclassical economics, where firms in an industry are identical, and respond similarly, quickly, and optimally to new information about the external environment. Modifications to stringent neoclassical models, however, can accommodate persistence in R&D or differences between firms in their R&D activities. The following overview briefly discusses some possible modifications and their limitations.

Two concepts from the empirical literature in industrial organization economics have provided explanations of persistence in R&D: adjustment costs and adaptive expectations. Adjustment costs stem from difficulties firms face in altering R&D expenditures in the short run. Such difficulties raise the cost to firms of quickly making large changes in R&D expenditures, and in turn cause firms to alter R&D more slowly than they would otherwise. Adjustment costs may arise from either capital market or labor market imperfections. Capital market imperfections (Himmelberg and Petersen 1991), caused by information asymmetries between firms and suppliers of external finance (Stiglitz and Weiss 1981, Myers and Majluf 1984), create a situation where firms who have superior information regarding future profits from R&D may be willing to fund R&D that outsiders will not (except perhaps at extraordinarily high costs to firms of external funds). Restrictions on external funding make it difficult for firms to expand R&D quickly if they lack sufficient internal funds. Adjustment costs also stem from labor market imperfections related to the frequent firing and rehiring of research personnel (Grabowski 1968). A firm will be reluctant to fire workers if they require firm-specific knowledge which makes training of new workers expensive (Himmelberg and Petersen 1991).

A positive correlation between current and past R&D expenditures can result not only from adjustment costs but also from adaptive expectations (Schankerman and Nadiri 1984). In the context of R&D funding decisions, adaptive expectations refer to firms' expectations of returns to R&D, and in particular to the expected price (of inputs for R&D) and quantity of output components of expected returns. Technically, a firm has adaptive expectations when its current expectation regarding the future amount or value of a specific economic variable is a weighted average of all past amounts or values of that variable. If a firm uses adaptive expectations of input prices and output levels to determine R&D expenditures, then current and past R&D activity will be correlated because current R&D activity is based on the history of past prices and quantities, much of which influenced past R&D activity as well.

Adaptive expectations and adjustment cost explanations of persistence in R&D, although not antithetical to evolutionary theory, do not lie at the core of the theory. In practice, both adaptive expectations and adjustment costs have been used primarily to modify what are otherwise optimizing models of firm investment in physical capital and R&D. In addition, neither adjustment costs nor adaptive expectations explicitly deal with differences between and within firms in their R&D activities.

Other modifications to a rigorous optimization model can produce differences between firms in equilibrium. A prototypic model by Lippman and Rumelt (1982), although it does not deal with R&D, shows that if firms have different qualities of initial endowments of assets and cannot perfectly replicate or find substitutes for other firms' assets (due to various sorts of uncertainty and imperfect information), then in equilibrium firms which maximize profits will have different levels of efficiency. Due to "substantial analytical difficulties" (Lippman and Rumelt 1982, p. 421), the Lippman and Rumelt model deals only with firms producing a single, homogeneous product, operating under conditions of a stable technology and a stable external environment. The model also does not explicitly address the issue of persistence. (In a completely stable environment with stable technology, of course, equilibrium efficiency differences will persist.) More generally, the maximization (or minimization) and equilibrium conditions required to simultaneously account for uncertainty, diversity, and change in formal economic models make it difficult for orthodox economics to deal with such complex phenomena (Nelson and Winter 1982, p. 28). While economic models have been constructed that explain *either* persistence in firm activities *or* differences between firms, formal economic models have much greater difficulty incorporating both aspects into one model. Conceptually such models are possible, of course. The numerous required modifications to neoclassical models and the difficulty of solving extremely complex orthodox models, however, suggests that a simpler explanation such as evolutionary theory may be more plausible.

This study does not aim to refute orthodox economics, a notoriously difficult undertaking. Instead, the research tests propositions which derive in a straightforward manner from evolutionary theory, and which require much greater complexity to derive from neoclassical economic models. As the introduction to this study noted, little prior research has systematically investigated persistent differences in firms' R&D activities. Some studies have found evidence of persistence in R&D. Rasmussen (1973) found a positive correlation between annual changes in firms' R&D expenditures and one-year lagged annual changes in firms' R&D expenditures; other studies have found a positive correlation between a measure of firms' current R&D capital stock (calculated per firm as the sum of several years past R&D expenditures, depreciated at an arbitrarily chosen rate) and one-year lagged capital stock (e.g., Nadiri and Bitros 1980). Such studies, however, do not also deal with differences between firms. Additional studies (Hall and Hayashi 1989, Himmelberg and Petersen 1991) have found that *within* firm variance of total R&D expenditures through time is relatively low, and that most of the variance in total firm R&D expenditures (after adjustment for firm size) is cross-sectional, reflecting differences *between* firms. These findings, however, are tangential to the main foci of these studies, and have not been pursued further.

Unlike much mainstream economic research, more behaviorally oriented economics (including that sympathetic to evolutionary theory) has drawn attention to and begun to document *within industry differences* in R&D policies. Researchers have investigated variation in whether or not firms undertake R&D, the amount of investment in R&D, and the emphasis on basic versus applied R&D. (See Dosi (1988) and Freeman (1982) for a comprehensive review and additional discussion.) By disaggregating R&D expenditures according to type of business application within the firm and tracking these disaggregated expenditures through time, this study will provide much more detailed information regarding persistent differences in R&D.

4. The Resource-based View

Although evolutionary theory diverges in significant ways from neoclassical economics, it has much more in common with resource-based theories in strategic management. The resource-based view emphasizes the heterogeneity between firms of their tangible and intangible assets and capabilities (Amit and Schoemaker 1992, Barney 1991, Mahoney and Pandian 1992, Peteraf 1993, Wernerfelt 1984). From the resource-based perspective, not only individual assets and capabilities but also combinations of resources contribute to a firm's competitive advantage (or disadvantage) (Conner 1991, Rumelt 1984). Recently, the resource-based view has begun to draw on evolutionary theory to address the dynamic aspects of building capabilities within the firm (Teece et al. 1994). The resource-based view contains the implication that if R&D within the firm is tied to particular assets which are heterogeneous and somewhat immobile across firms (e.g., oil recovery R&D geared to extraction of oil from individual geologic formations, no two of which are alike), and if possession of these complementary and heterogeneous assets is path dependent, then persistent differences in R&D might result not only from the characteristics of R&D itself, but also as a result of ties to heterogeneous assets.

The foregoing logic implies that arguments from evolutionary theory regarding firms' knowledge bases apply to knowledge related to the use of physical assets as well. An in-depth examination of the evolution of bundles of resources within firms is beyond the scope of this paper. To take into account the possibility that ties to firm assets determine R&D activity, however, the empirical analysis will incorporate assets related to specific R&D activities as possible explanatory variables for firms' R&D expenditures.

5. Petroleum Firm Data

The empirical investigation here focuses on R&D by major U.S. energy firms. The data come from the Financial Reporting System (FRS) of the U.S. Department of Energy (DOE). The FRS data base includes 26 major energy companies, chosen by the DOE from the top fifty publicly-owned domestic crude oil producers. Each firm in the FRS data base had at least one per cent of either the production or reserves of oil, gas, coal, or uranium; or one per cent of oil production, refining capacity, or petroleum product sales in 1976 (U.S. DOE 1982). Essentially, the sample includes the major domestic petroleum producers and refiners, plus a couple of railroad companies with large holdings of oil, gas, and coal reserves.[1] The FRS data base contains unique annual R&D expenditure data from 1974 to the present, broken down by types of R&D activity (e.g., oil and gas recovery, oil shale) within the firm. This detailed a firm-level breakdown of R&D expenditures over this long a time period for any set of firms is rare.[2] In addition, the data include a similar firm-level breakdown of balance sheet data such as assets and sales revenues.

The empirical analysis undertaken here utilizes R&D expenditures as a proxy for R&D effort. The expenditure levels for the various R&D activities provide a measure of the resources devoted to the R&D efforts, and also reflect company policies for allocation of R&D funds. As noted in the discussion of Hypothesis 1, both the levels of R&D effort and the allocation policies for R&D funds may persist through time. Because the large petroleum company mergers that occurred from 1982 through 1984 make it difficult to trace the R&D expenditures of the firms that merged, the analysis terminates in 1981. The resulting eight years of data from 1974 through 1982 provide a long enough time period over which to examine at least the medium-term stability or instability and similarities or differences of these firms' R&D activities.

This study examines five types of R&D undertaken by the FRS companies: oil and gas recovery, refining, coal gasification / liquefaction, oil shale, and tar sands. Only refining involves significant R&D on new products (e.g., gasoline additives, lubricants); new refined oil products tend to be proprietary (and, therefore, firm-specific). The other R&D activities, as well as a portion of refining R&D, involves research on processes. The research activities include efforts to extract and process oil shale and tar sands into crude oil substitutes; to process coal into refined oil or natural gas substitutes; to locate new crude oil deposits; to extract additional oil from previously drilled wells; to reduce costs of oil refining; and to process higher sulfur crude oil in existing refineries. Although the underlying technologies for the five R&D activities differ in the extent of their development, most of the R&D is applied and developmental (rather than basic research). This study excludes some other types of R&D by petroleum companies, namely R&D related to petrochemicals, coal mining, and nuclear power, for which firms outside of the petroleum industry performed a good deal of the R&D.

To provide a snapshot of the R&D data, for each of the five R&D activities and for total firm R&D, Tables 1 and 2 report the number of firms having positive R&D expenditures and total R&D expenditures for the FRS companies for each year 1974 through 1981. As Table 1 indicates, 24 out of 26 of the companies undertook some R&D during at least one year in the sample. Of

[1] The sample may have a selection bias. The data base includes only very large crude oil producers, and within that set, includes companies with large reserves or production of the main energy products in the U.S. This study's results, therefore, clearly are industry-specific and pertain only to these major, large energy companies. The companies in the sample are: Amerada Hess, American Petrofina, Ashland, Atlantic Richfield, Burlington Northern, Cities Service, Coastal Corp., Conoco, Exxon, Getty Oil, Gulf Oil, Kerr-McGee, Marathon Oil, Mobil, Occidental, Phillips Petroleum, Shell Oil, Standard Oil of California, Standard Oil Company (Indiana), Standard Oil Company (Ohio), Sun Company, Superior, Tenneco, Texaco, Union Oil of California, and Union Pacific.

[2] This R&D breakdown is not publicly available, in annual reports or elsewhere.

Table 1 **Number of Firms with Positive R&D Expenditures (Total Number of Firms in Sample = 26)**

	R&D Activity					
Year	Oil & Gas Recovery	Other Petroleum (Refining)	Coal Gasification/ Liquefaction	Oil Shale	Tar Sands	Total R&D*
1974	17	18	14	11	6	22
1975	17	18	14	13	8	22
1976	17	17	13	12	9	22
1977	18	17	14	15	10	22
1978	16	16	14	16	11	21
1979	17	18	14	15	12	22
1980	16	19	17	15	12	23
1981	17	17	18	17	12	24

* Includes nonenergy R&D and miscellaneous other energy R&D.

Source: Financial Reporting System, Energy Information Administration, U.S. Department of Energy.

the five R&D activities examined here, oil and gas recovery and refining had the greatest number of firms conducting R&D during the period, followed by coal gasification / liquefaction, oil shale, and tar sands R&D. The relative number of firms in each R&D activity roughly accords with the relative development of the associated technologies: oil and gas extraction and refining are the most developed technologies, followed by coal gasification / liquefaction (some first-generation, low btu, coal gasification and liquefaction plants have been used in other countries), and then followed by oil shale and tar sands-based technologies. R&D expenditures shown in Table 2 also follow this pattern, with the greatest expenditures on the most developed technologies and the least expenditure on the less developed technologies.

Conditions in the external selection environment affected R&D expenditures in the oil industry during the

Table 2 **Industry Research and Development Expenditures (Million Dollars—Current Dollars)**

	R&D Activity					
Year	Oil & Gas Recovery	Other Petroleum (Refining)	Coal Gasification/ Liquefaction	Oil Shale	Tar Sands	Total R&D*
1974	168	189	42	12	18	702
1975	191	184	36	14	8	794
1976	210	187	43	12	9	907
1977	209	219	82	10	13	1,021
1978	250	254	146	9	17	1,283
1979	339	273	175	13	28	1,566
1980	422	359	188	23	32	2,034
1981	668	468	268	92	61	2,820
TOTAL	2,457	2,133	980	185	186	11,127

* Includes nonenergy R&D and miscellaneous other energy R&D.

Source: Financial Reporting System, Energy Information Administration, U.S. Department of Energy.

1970s. Table 2 shows that expenditures for all five R&D activities rose from 1974 through 1981. During that period rising oil prices not only made more funds available for R&D as profits rose, but also rising prices increased the expected returns to R&D. All five R&D activities relate to the production or sale of crude or refined oil products or their substitutes, the prices for all of which increased.[3] As expected returns to the less proven technologies increased, some firms not previously engaged in these activities (particularly oil shale and tar sands R&D) began conducting R&D, as shown in Table 1. Such R&D entry in the less proven technologies could reduce the observed degree of path dependence in these R&D activities. This R&D entry, however, probably does not fully explain any differences in path dependencies across the five R&D activities. In particular, not all firms began conducting R&D in all of the less proven technologies, suggesting that the extent of path dependence in these technologies may reflect other factors in addition to R&D entry driven by changes in the external environment. Such factors include internal firm influences on patterns of R&D activity.

6. Empirical Evidence

Much empirical research on R&D expenditures has focused on the firm's R&D to sales ratio (total firm R&D expenditures divided by total firm sales revenues), also termed "R&D intensity." It is well-established that within an industry, firm R&D expenditures rise approximately linearly with sales revenues (Scherer 1992). The use of an R&D intensity variable allows researchers to directly control for an effect of firm size on R&D expenditures. Since the hypotheses examined here do not involve firm size, to control for the likelihood that larger firms spend more on R&D than smaller firms, annual firm expenditures for each R&D activity are divided by annual sales revenue for the entire firm.[4]

The use of R&D intensity variables in the analysis raises the obvious question of how these variables relate to the earlier hypotheses about firms' R&D activities. With regard to differences between firms, the R&D intensity variables have the advantage of controlling for a large source of differences between firms that has little to do with the arguments made earlier. With regard to persistence in R&D, the use of R&D intensity could understate the degree of path dependence in R&D if sales revenues fluctuate more than do R&D expenditures. Alternatively, R&D intensity could vary less through time than do R&D expenditures if firms' budgetary decision rules keep R&D expenditures roughly proportional to sales: R&D expenditures would fluctuate but R&D intensity would not. Under this scenario, the use of R&D intensity in the analysis could overstate path dependence relative to actual R&D expenditures. Regardless of whether R&D intensity may overestimate or underestimate the path dependence of R&D expenditures, however, the use of an intensity variable has the advantage of controlling for the impact on R&D of changes in firm size over time that are not related to the earlier hypotheses.

Table 3 provides simple statistics for the R&D intensity variables for all five R&D activities, and also includes simple statistics for an annual dummy variable coded "1" if the firm had positive R&D expenditures in a given activity and "0" if not. In what follows, these variables are analyzed for evidence of path dependencies, between firm differences, and within firm differences in R&D activities. Correlation coefficients provide preliminary evidence regarding path dependence in each R&D activity (Hypothesis 1), and coefficients of variation of the R&D intensity variables provide evidence regarding differences between and within firms in their R&D ac-

[3] In the case of refinery operations, the price of the major input (crude oil) also rose. Anticipated returns to R&D related to refinery operations and refined products therefore may not have risen to the same extent as for the other R&D activities.

[4] An alternative procedure of dividing R&D expenditures for each activity by sales revenues related just to firm operations in that activity (e.g., the use of sales revenues from oil shale operations as a divisor for R&D expenditures on oil shale) is not possible for all five R&D activities using the FRS data. The data base does not contain a breakdown of sales revenues for coal gasification/liquefaction, oil shale, and tar sands. Such a procedure also would not control for firm size, but would only control for amount of sales in an operation related to each R&D activity. Another procedure to account for firm size might involve including sales revenue as one of several independent variables in a regression where R&D spending in a particular activity is the dependent variable. Unfortunately, this poses practical problems of multicollinearity, because sales revenues are highly correlated with many of the right-hand side variables to be used in the regression analysis here.

Table 3 R&D Variables Simple Statistics 1974-81 (208 observations per R&D activity)

Firm Expenditures per R&D Activity/Total Firm Sales Revenue

R&D Activity	Mean	Standard Deviation
Oil & Gas Recovery	0.000688	0.000839
Refining	0.000558	0.000635
Coal Gasification/ Liquefaction	0.000180	0.000366
Oil Shale	0.000195	0.000889
Tar Sands	0.0000568	0.000182
Total Firm R&D	0.00335	0.00251

Dummy Variable Indicating Positive R&D Activity Expenditures (1 = positive firm R&D expenditures, 0 = zero firm R&D expenditures)

R&D Activity	Mean	Standard Deviation
Oil & Gas Recovery	0.647	0.478
Refining	0.673	0.470
Coal Gasification/ Liquefaction	0.567	0.497
Oil Shale	0.548	0.499
Tar Sands	0.385	0.488
Total Firm R&D	0.856	0.352

Source: Financial Reporting System, Energy Information Administration, U.S. Department of Energy.

tivities (Hypotheses 3 and 4). Regression analysis of each R&D activity provides additional evidence regarding path dependence (Hypothesis 1), within firm differences across R&D activities (Hypothesis 4), and incremental change in R&D expenditures (Hypothesis 2). In the following discussion of the empirical results, many of the footnotes deal with technical aspects and possible limitations of the statistical techniques employed.

6.1. Correlation Coefficients

To provide preliminary evidence regarding Hypothesis 1 of path dependence in R&D, for each R&D activity, for both the R&D intensity and R&D dummy variables, Table 4 reports Pearson correlation coefficients of current year observations with observations one to seven years prior. The following procedure was used to compute these correlations. For each type of R&D and each type of variable (intensity or dummy), as many observations in the sample as possible were lagged one, two, three, four, five, six, and seven years. (Only observations from 1981 could be lagged for all of these lengths of time; observations from other years have fewer year lags, since the data begin only in 1974.) Then for each R&D variable (two per type of R&D), seven subsamples were created. Each subsample contains the original observations and the observations lagged either one, two, three, four, five, six, or seven years. (The resulting subsamples have different numbers of observations, since not all observations could be lagged for all numbers of years.) In each subsample, the original and lagged observations were paired by firm, and correlation coefficients between the current and lagged observations were computed.[5] Thus, for each R&D activity variable, Table 4 contains seven correlation coefficients, one each for a given number of years (one through seven) between the firm-level observations.

In Table 4, a comparison across R&D activities of the correlation coefficients of current and one-year lagged share variables yields the following ranking of coefficients from highest to lowest: oil and gas recovery, refining, oil shale, coal gasification / liquefaction, and tar sands. The correlation coefficients for all but tar sands R&D are very high—0.80 or above—indicative of path dependence as suggested by Hypothesis 1. The correlation coefficients of current and one year-lagged R&D dummy variables are above 0.80 for all five R&D activities. As the time lag between observations increases, the correlation coefficients for both the R&D intensity and dummy variables tend to fall for all of the R&D activities (although oil shale R&D intensity has a seemingly anomalous high correlation of current observations and those lagged seven years).[6]

To serve as a basis for comparison with the R&D activity correlations, similar lagged correlation coefficients

[5] These calculations pool all firms in the sample. If instead the correlation coefficients for individual firms differ from one another, the reported correlations will have some bias. Some rough simulations (available on request) suggest that any bias would be unlikely to change the *comparative* results between types of R&D and number of year lags.

[6] Spearman correlation coefficients (of annual numerical rankings within the FRS sample of the R&D intensity variables) for current observations with observations one to seven years prior provide similar evidence of path dependence to that presented in Table 4. This information is available on request from the author.

Table 4 Correlation Coefficients of Current and Lagged R&D Expenditure Variables (1974–81)

	Pearson Coefficients* (Figures Rounded to Nearest Thousandth)						
R&D Activity	1 Year Lag (182 obs.)	2 Year Lag (156 obs.)	3 Year Lag (130 obs.)	4 Year Lag (104 obs.)	5 Year Lag (78 obs.)	6 Year Lag (52 obs.)	7 Year Lag (26 obs.)
Oil & Gas Recovery							
R&D $/Sales $	0.940	0.921	0.886	0.853	0.863	0.849	0.818
Positive R&D Dummy	0.904	0.901	0.848	0.854	0.860	0.875	0.830
Refining							
R&D $/Sales $	0.922	0.852	0.811	0.786	0.789	0.740	0.631
Positive R&D Dummy	0.913	0.827	0.741	0.694	0.673	0.639	0.566
Coal Gasif./Liquef.							
R&D $/Sales $	0.808	0.632	0.634	0.577	0.383	0.365	0.557
Positive R&D Dummy	0.867	0.835	0.820	0.736	0.704	0.588	0.553
Oil Shale							
R&D $/Sales $	0.901	0.790	0.691	0.730	0.619	0.505	0.855
Positive R&D Dummy	0.824	0.720	0.685	0.634	0.489	0.494	0.459
Tar Sands							
R&D $/Sales $	0.405	0.145	0.294	0.322	0.438	0.151*	0.166*
Positive R&D Dummy	0.864	0.739	0.613	0.543	0.473	0.482	0.408

* All coefficients significant at the 0.05 level except those marked with an asterisk.

were computed for categories of investment related as closely as possible to the five R&D activities. Table 5 reports the Pearson correlation coefficients for annual investment expenditures in relevant investment categories divided by total annual firm sales revenue. With the exception of investment in petroleum production, the very low correlation coefficients for the investment activities suggest low path dependence. And although

Table 5 Correlation Coefficients of Current and Lagged Investment Expenditure Variables (1975–81)**

	Pearson Coefficients* (Figures Rounded to Nearest Thousandth)					
R&D Activity	1 Year Lag (156 obs.)	2 Year Lag (130 obs.)	3 Year Lag (104 obs.)	4 Year Lag (78 obs.)	5 Year Lag (52 obs.)	6 Year Lag (26 obs.)
Petroleum Production						
Invest. $/Sales $	0.715	0.558	0.450	0.258	0.404	0.453
Refining						
Invest. $/Sales $	0.087*	0.118*	0.334	0.032*	0.070*	Not Available
Coal Mining & Gasif./Liquef.						
Invest. $/Sales $	0.020*	−0.336	0.032*	0.004*	0.151*	Not Available
Oil Shale						
Invest. $/Sales $	0.167	−0.014*	−0.003*	0.043*	0.117*	0.015*
Tar Sands						
Invest. $/Sales $	0.579	0.342	−0.183*	−0.479	−0.485	−0.716

* All coefficients significant at the 0.05 level except those marked with an asterisk.

** Investment data are not available for 1974.

Source: Financial Reporting System, Energy Information Administration, U.S. Department of Energy.

the correlation coefficients for the investment variables related to petroleum production do indicate some path dependence, the correlation coefficients are much lower than for the oil recovery R&D variables. These figures suggest that all five of the R&D activities have greater stability over time than another firm-level activity—that of investment—with which R&D expenditures are often thought to be associated. This stability of R&D relative to investment provides additional support for hypothesis 1 that R&D activities in general exhibit path dependence.

6.2. Variation Across Firms and R&D Activities

In addition to the path dependence of R&D, Hypotheses 3 and 4 advanced earlier deal with differences between and within firms in their R&D activities. To provide evidence regarding these hypotheses, for each R&D activity, each firm's annual R&D intensity variables are averaged over the period 1974–81. This provides a summary measure of expenditures per firm and per R&D activity. The top half of Table 6 reports the mean, standard deviation, and coefficient of variation of the firm-level averages for each R&D activity. The coefficient of variation (standard deviation divided by the absolute value of the mean) for each variable yields a standardized measure of differences in R&D effort between firms in each R&D activity (the larger the coefficient of variation, the greater the differences), the focus of Hypothesis 3.

The coefficients of variation for the five R&D activities in the top half of Table 6 differ a good deal and also substantially exceed the coefficient of variation for the firm-level average of total firm R&D (used as a benchmark). This dispersion of firm expenditures in the individual R&D activities provides support for Hypothesis 3 that in a given R&D activity firms differ in their levels of effort. Not surprisingly, R&D expenditures in the less developed technologies show greater variation across firms, perhaps as a result of greater uncertainty (and therefore less consensus) regarding the expected returns to such R&D.

By construction, the firm-level averages in each R&D activity mask changes in R&D intensity over time, which could contribute to the divergent firm-level averages. In order to compare the differences across firms with changes through time, for each R&D activity an average of the 26 FRS firms' R&D intensities is computed for each year 1974 through 1981. The mean, standard deviation, and coefficient of variation of these annual industry averages for each R&D activity are then computed. This information is presented in the lower half of Table 6. If the differences in R&D intensity between firms are greater than the differences in industry R&D intensity over time, then in each R&D activity the coefficient of variation for the firm-level average should exceed the coefficient of variation for the annual industry average. Table 6 shows exactly this result. For each R&D activity, the coefficient of variation for the firm-level average is several times greater than the coefficient of variation for the annual industry average. Differences between firms exceed differences through time in

Table 6 Evidence Regarding Between Firm Variation in R&D Activities[1]

	Firm Average R&D Intensity 1974–81[2] (26 observations per R&D activity)		
R&D Activity	Mean	Standard Deviation	Coefficient of Variation
Oil and Gas Recovery	0.000688	0.000811	1.179
Refining	0.000558	0.000592	1.061
Coal Gasif./Liquef.	0.000180	0.000302	1.678
Oil Shale	0.000195	0.000707	3.626
Tar Sands	0.000057	0.000112	1.965
Total Firm R&D	0.003349	0.002326	0.696
	Annual Industry Average R&D Intensity[3] (8 observations per R&D activity)		
R&D Activity	Mean	Standard Deviation	Coefficient of Variation
Oil and Gas Recovery	0.000688	0.000079	0.115
Refining	0.000558	0.000071	0.128
Coal Gasif./Liquef.	0.000180	0.000059	0.327
Oil Shale	0.000195	0.000120	0.612
Tar Sands	0.000057	0.000030	0.534
Total Firm R&D	0.003349	0.000205	0.061

[1] Figures rounded to the nearest millionth. Coefficients of variation rounded to nearest thousandth.

[2] Firm average over the period 1974–81 of annual R&D expenditures per activity divided by total firm sales for each R&D activity.

[3] Industry average R&D intensity per year (1974 through 1981), for each R&D activity.

an R&D activity, providing strong support for Hypothesis 3.

In addition to Hypothesis 3 regarding differences between firms in a given R&D activity, Hypothesis 4 deals with differences within and between firms across R&D activities. Although Table 6 shows that in each separate R&D activity firms differ from one another in their effort levels (consistent with Hypothesis 3), this could arise from each individual firm spending equally on all five R&D activities. Further examination of the evidence on Table 6, however, sheds doubt on this possibility. In particular, in the top half of Table 6 the mean values of the firm-level averages differ substantially by type of R&D. These differences in means indicate that the FRS firms did not divide R&D funds equally among R&D activities. If instead each individual firm had allocated funds equally between R&D activities, all R&D activities would have the same means for the firm-level averages. Clearly, the data support the first part of Hypothesis 4, which states that an individual firm does not devote the same degree of effort to all R&D activities.

The second part of Hypothesis 4 suggests that the pattern of within firm differences in R&D also differs between firms. To examine this aspect of firm differences, a new variable is created that directly incorporates differences between firms in an individual R&D activity. This variable is the firm's annual share of industry (FRS company) annual R&D activity expenditures divided by the firm's share of industry annual sales revenue. Such an R&D "share" variable captures relative differences between firms' R&D expenditures, adjusted for differences in overall firm size. For each R&D activity, to provide a summary measure per firm, each firm's annual "share" variables are averaged over the period 1974–81. Each firm has five average share variables, one per R&D activity.

By construction, the firm-level average share variables make it possible to investigate differences between firms in the pattern of within firm differences in the allocation of R&D effort across activities. If each firm allocates R&D funds to each of the five R&D activities in the same proportion to total firm R&D as do all other firms in the industry, then each firm's share of industry spending in an activity would be the same for all R&D activities, as would each firm's size-adjusted shares. Therefore, the dispersion of the average (from 1974–81) size-adjusted R&D shares across firms in each R&D activity would be the same for all R&D activities. The results in Table 7, however, counter this supposition. The coefficients of variation of the share variables differ a good deal between R&D activities, providing support for the contention in Hypothesis 4 that the pattern of within firm differences in R&D activities differs between firms as well.

The evidence presented thus far provides initial support for Hypotheses 1, 3, and 4. The positive autocorrelations of the R&D intensities provide support for Hypothesis 1 that past and present R&D activity is positively correlated. The coefficients of variation of the firm average R&D intensities provide support for Hypothesis 3 that in each R&D activity, the level of effort differs across firms. In addition, differences in the means of the firm average R&D intensities support the first proposition of Hypothesis 4 that firms devote different degrees of effort to different R&D activities. And differences in the coefficients of variation of the average R&D share variables support the second proposition of Hypothesis 4 that the pattern of within firm differences in R&D effort differs across firms. Although these statistics support Hypothesis 1, 3, and 4, the tests do not control for influences on R&D activity unrelated to the hypotheses except for firm size. Regression analysis can control for additional influences on R&D activity, and therefore provide more fine-grained tests.

Table 7 Evidence Regarding Within Firm Variation in R&D Activities (figures rounded to the nearest thousandth)

	Firm Average R&D Share 1974–81[4] (26 observations per R&D activity)		
R&D Activity	Mean	Standard Deviation	Coefficient of Variation
Oil and Gas Recovery	0.785	0.929	1.183
Refining/Marketing	0.695	0.740	1.065
Coal Gasif./Liquef.	0.575	0.999	1.737
Oil Shale	3.750	14.007	3.735
Tar Sands	0.861	1.673	1.943
Total Firm R&D	0.848	0.589	0.695

[4] Average per firm over the period 1974–81 of the firm's share of industry R&D expenditure per R&D activity divided by the firm's share of total industry sales revenues.

6.3. Regression Analysis

Separate regressions for each R&D activity, as well as a regression which pools R&D activities, further investigate path dependence (Hypothesis 1) and within firm variations in R&D effort (Hypothesis 4). In addition, a second set of regressions provides evidence regarding incremental change in firms' R&D expenditures over time (Hypothesis 2).

6.3.1. Independent Variables. In the first set of separate regressions for each R&D activity, the dependent variable in each regression is R&D intensity for the particular activity. The independent variables used in each of the regressions are:

(1) lagged (one year) dependent variable;

(2) total firm R&D expenditures minus the firm's expenditures for the type of R&D activity used as the dependent variable, adjusted for firm size (by dividing by the firm's sales revenue);

(3) investment (annual change in property, plant, and equipment) in assets related to the R&D activity (e.g., mineral deposits or plant and equipment to which the R&D applies), adjusted for firm size;

(4) value of related assets adjusted for firm size, lagged one year;[7]

(5) sum over all FRS firms of the squared individual firm shares of industry R&D expenditures for the R&D activity used as the dependent variable, minus the firm's own squared share of industry expenditures;

(6) dummy variable for each year, excluding 1981.

The rationale for the inclusion of each of the independent variables in the regressions is as follows. The influence of the lagged dependent variable provides a measure of path dependence. The higher the estimated coefficient on the lagged dependent variable, the greater the path dependence.[8] The second right-hand side variable, total firm R&D expenditures minus firm expenditures on the R&D activity in question (and divided by firm sales revenue), is used to measure the similarity of a firm's expenditures on the R&D activity in question with expenditures on the rest of the firm's R&D. Two additional independent variables control for the possibility, suggested by the resource-based view of the firm, that relationships between R&D and related firm resources influence R&D activity expenditures. The investment variable accounts for the possibility that R&D expenditure patterns could primarily reflect investment patterns if firms gear their R&D expenditures to their investment programs; the lagged asset variable accounts for the possibility that past acquisition of assets may lock firms into related R&D activity. (Because FRS asset data are year-end values, the lagged asset variable reflects the stock of assets at the start of each year.) The sum of squared R&D expenditure shares controls for a possible effect on R&D expenditures of rivalry in R&D between firms. The rivalry measure is essentially a herfindahl index for R&D expenditure shares, excluding each firm's own squared R&D share. The exclusion of own-firm squared R&D expenditures permits the herfindahl measure to capture the impact on each firm only of other firms' spending; the squaring of other firms' expenditure shares captures the likelihood that competition from the larger spenders will have the greatest effect on own-firm R&D spending.

The last control variables are dummy variables for each year included in the regressions through 1980. Because the use of dummy variables for individual years requires that the regression exclude a dummy variable for one year, the dummy variable for 1981, the last year in the sample, is excluded. (The constant term will estimate the effect of that year on the dependent variable.) The year dummy variables control for external environmental factors that might have affected R&D expenditures, such as energy prices or government programs (e.g., the proposed Synfuels program in the late 1970s and early 1980s, and synfuels R&D subsidies provided by the government in the 1970s). Such factors may have increased R&D expenditures: subsidies for research could have raised total expenditures (including the subsidies) on synthetic fuels R&D, and higher energy prices increased the expected profitability of the energy products that might result from R&D. Because energy prices probably affected many of the variables in the regression, including investment spending as well as the sales revenue denominator used in constructing all but the rivalry and dummy variables, energy prices

[7] For refining and coal gasification/liquefaction R&D, missing data in 1974 shorten the sample by one year.

[8] This logic applies to coefficient estimates of zero or greater. This study also does not attempt to piece out the lag structure of path dependence. As a simple approximation, a one-year lag should indicate whether or not path dependence is present.

are not included directly as separate regressors. Instead, the year dummy variables should pick up any response of the dependent variable to at least any major shift in energy prices not reflected by other variables in the regression.

The control variables raise some econometric issues. First, the investment and asset variables are positively correlated and significant at the 0.05 level or less for oil and gas recovery (correlation coefficient = 0.75, $p < 0.01$) and for tar sands (correlation coefficient = 0.26, $p < 0.01$).[9] Since the investment and asset variables capture somewhat different possible influences on R&D expenditures, however, both variables separately are included in the regressions. Tests of joint significance of the investment and asset variables are used in the regressions where both variables are insignificant. Second, the issue of simultaneity arises with regard to R&D and investment: R&D may influence investment spending, as well as the reverse. In the petroleum industry, however, it is unlikely that investment spending depends significantly on R&D undertaken in the same year. Although the industry as a whole spends large amounts on R&D, the relatively low mean total firm R&D to sales ratio of 0.33 percent (reported in Table 3) indicates that R&D is unlikely to drive operations and investment to the same extent as it does in more R&D intensive industries such as pharmaceuticals and semiconductors.

The rivalry variable used in the regressions also merits discussion. Unlike the other non-dummy variables, the rivalry variable for each R&D activity does not include an adjustment for firm size, primarily because competition in R&D involves all of firms' expenditures for a given R&D activity, not just that portion not accounted for by firm size. The rivalry variable by construction may have a negative relationship with own-firm R&D spending (the numerator of the dependent variable in each regression): although the rivalry variable for each firm excludes that firm's own squared R&D expenditure share, the squared expenditure shares for other firms include own-firm R&D in the denominator (i.e., in total FRS firm expenditures that form the denominator of the R&D expenditure shares). The obvious alternative of excluding the firm's own R&D from the denominator, however, could produce highly misleading measures of the rivalry faced by each firm. Consider a hypothetical example of a firm with an 80 percent expenditure share and only one rival with positive R&D spending: A rivalry measure that excludes own-firm R&D from the denominator would show the rival as having a monopoly on R&D spending.

The use of the rivalry variable requires two final caveats. First, the analysis does not attempt to account for possible simultaneity in the relationship between R&D intensity and the expenditures of other firms. Secondly, the relationship between R&D intensity and the market structure of the industry (e.g., concentration in industry sales), the focus of much work in economics, is not incorporated. Because the petroleum industry is not R&D driven, the linkage between market structure and R&D is unlikely to be strong.

6.3.2. Initial Regression Results. Because the dependent variable for each of the R&D activities has some observations with zero values (i.e., the observations for each of the dependent variables are censored), the regressions use Tobit analysis.[10] Table 8 reports an initial set of tobit regression results for each R&D activity.[11] In addition, in order to make comparisons of the estimated coefficients within and across equations, Table 8 reports "standardized" coefficient estimates for the Tobit regressions. For each right-hand side variable in each regression, the standardized coefficient estimate equals the original estimated Tobit coefficient multiplied by the standard deviation of the variable and divided by the estimated standard deviation of the uncensored de-

[9] Correlation matricies and summary statistics for the independent variables in the regressions are not included here due to space constraints. This information is available on request from the author.

[10] The only software available for use with the FRS database is SAS version 5. Because SAS version 5 does not have a tobit procedure, the estimation uses a program in SAS matrix language (adapted from one in a SAS User's Guide) to compute tobit maximum likelihood estimates using Fair's technique (Fair 1977).

[11] The pooling of time periods in the panel may mask changes in the coefficients through time. In addition, the pooling of firms may bias the coefficients if individual firms actually have different coefficients (Theil 1971). In regressions run for individual firms with only the lagged dependent variable on the right hand side, however, the coefficients proved unstable. Lack of degrees of freedom due to the short time series prevent the use of regressions for individual firms that include all four independent variables.

pendent variable.[12] Each standardized coefficient measures how large a change in the standard deviation of the dependent variable results from a one standard deviation change in the right-hand side variable. Table 8 also reports OLS results. Except where noted, the following discussion focuses on the Tobit regressions.

The results of these initial regressions confirm the previous analyses using correlation coefficients and coefficients of variation. For all five R&D activities, the coefficient on the lagged R&D intensity variable is positive and statistically significant at the 0.01 level. This result supports Hypothesis 1 that current and past R&D activities are positively correlated. In addition, all of the R&D activities except tar sands have high path dependence: the standardized Tobit coefficients on the lagged dependent variables range from 0.71 to 0.91 for the four R&D activities other than tar sands.

The regressions also provide support for Hypothesis 4 that firms do not devote the same effort to all R&D activities. Although for each of the five R&D activities the coefficient on the total firm R&D minus R&D activity expenditure variable is positive and differs significantly from zero, the standardized coefficients vary a good deal between regressions (ranging from 0.05 to 0.26). Thus, a one standard deviation increase in "total" firm R&D intensity is associated with differing proportionate increases in the R&D intensities of the five R&D activities. If firms devote the same effort to all R&D activities, however, the relationship between a change in R&D intensity for an R&D activity and a change in the remainder of the firm's R&D should be the same for all R&D activities.[13] The regressions show this is not the case.

The estimated coefficients on the rivalry variables differ significantly from zero, and the standardized coefficients for these variables have the next highest magnitudes in absolute value after those for lagged R&D intensity. Only for tar sands, however, does the standardized coefficient on the rivalry index exceed that for lagged R&D intensity in magnitude. Finally, of the investment and lagged asset variables, only the tar sands asset variable is individually significant. The investment and asset variables are jointly significant for each of the other four R&D activities, with standardized coefficient estimates ranging from −0.09 to 0.07. The lagged dependent variable, however, may pick up much of any impact of the lagged assets. The significant but low values of these coefficients provide limited support for the possibility, based on the resource-based view, that ties to related firm resources affect R&D activity.[14]

The residuals from the regressions reported in Table 8 were analyzed for evidence of heteroscedasticity and serial correlation. Plots of the residuals do not provide any evidence of heteroscedasticity. Estimates of first-order autocorrelation coefficients for the residuals in each regression, however, do suggest that serial correlation is present in some of the regressions.[15] Serial cor-

[12] In OLS regression with uncensored data, the divisor is the standard deviation of the dependent variable. In Tobit regression, the data are censored. Therefore, it is necessary to estimate the standard deviation of the uncensored dependent variable, and use this as the divisor. This can be obtained from the original Tobit regression, as follows. In matrix form, the regression is $Y = XB + E$, and Var (Y) = Var (XB) + Var (E), since XB and E are independent. The estimated variance of the uncensored dependent variable can be computed by using the estimated beta coefficients $(\hat{B})$ to obtain $X\hat{B}$ and Var $(X\hat{B})$, and adding Var $(X\hat{B})$ to the variance of the error term (Var E).

[13] To see this, suppose a firm has $g + 1$ different R&D activities, and spends X dollars on each activity. A one standard deviation increase in spending per R&D activity equals $SD(X)$, and a one standard deviation increase in the remainder of the firm's R&D spending equals $gSD(X)$. If firms spend equally (and thus change spending equally) for all R&D activities, then an increase of $gSD(X)$ on the "total" R&D variable in each regression should produce the same standard deviation increase in the dependent variable in all five regressions. This does not occur.

[14] The negative values of a few of the coefficients on the investment and asset variables run counter to the resource-based view, which would predict positive coefficients if firms undertake R&D which complements current investment and asset stocks. These negative coefficients individually also were insignificant in the regressions, however.

[15] To test for first-order serial correlation in each regression, the residuals for each firm separately were lagged one year. The current and lagged residuals were pooled for all firms, and OLS regression was used to estimate the autocorrelation coefficient. This procedure was applied to the residuals from each of original regressions (OLS and Tobit). For the Tobit regressions, the absolute values of all of the estimated autocorrelation coefficients were above 0.12 except that for the oil and gas recovery R&D residuals. For the OLS regressions, the estimated autocorrelation coefficients had absolute values above 0.12 except for the refining and oil shale residuals.

Table 8 R&D Intensity Regressions—Individual R&D Activities[a] (t-statistics in parentheses)

Dependent Variable: Oil & Gas Recovery R&D $/Total Firm Sales $

Regression Technique	Constant Term[b]	Lagged Dependent Variable	Petroleum Production Invest. $/Sales $	Lagged Petroleum Prod. Asset $/ Sales $	Total Firm R&D Minus Oil & Gas R&D $/ Sales $	Index of Rivalry in Oil & Gas R&D	Year Dummy Variables[b]	R^2	Adj. R^2	χ^2 Statistic	N
Tobit		1.0823*	−0.0008†	0.0002†	0.0233***	−0.0038***	*			1734.179*	182
		(31.064)	(−1.500)	(1.598)	(1.775)***	(−1.918)					
Standardized Tobit		0.9065	−0.0720	0.0692	0.0531	−0.0592					
OLS		0.9660*	−0.0002	0.0001	0.0043	−0.0032**	*	0.8999	0.8934		182
		(36.219)	(−0.611)	(0.656)	(0.441)	(−2.003)					

Dependent Variable: Refining R&D $/Total Firm Sales $

Regression Technique	Constant Term[b]	Lagged Dependent Variable	Refining Invest. $/Sales $	Lagged Ref./Mktg Asset $/Sales $	Total Firm R&D Minus Refining R&D $/Sales $	Index of Rivalry in Refining R&D	Year Dummy Variables[b]	R^2	Adj. R^2	χ^2 Statistic	N
Tobit		0.9700*	0.0028†	0.0006†	0.0390*	−0.0026**	***			1140.760*	156
		(20.621)	(1.504)	(1.183)	(2.928)	(−2.125)					
Standardized Tobit		0.8290	0.0513	0.0476	0.1175	−0.0890					
OLS		0.8722*	0.0024	0.0001	0.0129	−0.0027**	**	0.8633	0.8539		156
		(23.600)	(1.650)	(0.407)	(1.287)	(−2.600)					

Dependent Variable: Coal Gas./Liquef. R&D $/Total Firm Sales $

Regression Technique	Constant Term[b]	Lagged Dependent Variable	Coal Mining and Gas./Liq. Invest. $/Sales $	Lagged Coal Mining and Gas./Liq. Asset$/Sales $	Total Firm R&D Minus Coal Gas R&D $/Sales $	Index of Rivalry in Coal Gas R&D	Year Dummy Variables[b]	R^2	Adj. R^2	χ^2 Statistic	N
Tobit		0.9852*	−0.0003†	−0.0014†	0.0220***	−0.0011**	**			270.792*	156
		(10.854)	(−0.157)	(−1.314)	(1.9215)	(−2.294)					
Standardized Tobit		0.7090	−0.0107	−0.0898	0.1073	−0.2343					
OLS		0.7717*	0.0006	0.0004	0.0037	−0.0014*	***	0.7068	0.6866		156
		(12.571)	(0.691)	(0.649)	(0.492)	(−3.983)					

Dependent Variable: Oil Shale R&D $/Total Firm Sales $

Regression Technique	Constant Term[b]	Lagged Dependent Variable	Oil Shale Invest. $/Sales $	Lagged Oil Shale Asset $/ Sales $	Total Firm R&D Minus Oil Shale R&D $/ Sales $	Index of Rivalry in Oil Shale R&D	Year Dummy Variables[b]	R^2	Adj. R^2	χ^2 Statistic	N
Tobit		0.7364*	0.0088†	0.0104†	0.0506*	−0.0088*	*			601.314*	182
		(19.387)	(0.360)	(0.584)	(3.153)	(−4.582)					
Standardized Tobit		0.7883	0.0140	0.0252	0.1408	−0.3438					
OLS		0.6930*	0.0157	−0.0097	0.0016	−0.0060*	*	0.8395	0.8291		182
		(23.938)	(0.880)	(−0.812)	(0.154)	(−4.104)					

Dependent Variable: Tar Sands R&D $/Total Firm Sales $

Regression Technique	Constant Term[b]	Lagged Dependent Variable	Tar Sands Invest. $/ Sales $	Lagged Tar Sands Asset $/ Sales $	Total Firm R&D Minus Tar Sands R&D $/ Sales $	Index of Rivalry in Tar Sands R&D	Year Dummy Variables[b]	R^2	Adj. R^2	χ^2 Statistic	N
Tobit		0.2937*	0.0040†	0.0035*†	0.0274*	−0.0025*	*			74.276*	182
		(3.236)	(1.034)	(3.133)	(3.425)	(−4.668)					
Standardized Tobit		0.2111	0.0679	0.2042	0.2604	−0.6158					
OLS		0.2177	−0.00004	0.0015**	0.0024	−0.0015*	*	0.3240	0.2803		182
		(4.050)	(−0.018)	(2.334)	(0.619)	(−4.946)					

[a] Regressions include years 1975–81, except for refining R&D and coal gasification/liquefaction R&D which include the years 1976–81 (due to missing data).

[b] Estimates of the constant term and of the year dummy variable coefficients are not reported separately. The significance level indicated for the year dummy variables refers to their joint significance in the regression.

* Significant at the 0.01 level or less.

** Significant at the 0.05 level or less.

*** Significant at the 0.10 level or less.

† Investment and lagged asset variables jointly significant at the 0.05 level or less.

Note: Tests for joint significance of investment and asset variables in the regression are χ^2 tests in the Tobit regressions and F-tests in the OLS regressions, and significance levels are based on one-tailed tests. The significance levels of the χ^2 statistics in the Tobit regressions also are based on one-tailed tests. All other significance levels are based on two-tailed tests.

relation poses special complications in regressions involving lagged dependent variables. Tobit estimation of regressions with lagged dependent variables and serial correlation of the residuals is, in the words of Maddala (1983, p. 187), "very complicated." Standard corrections for serial correlation in OLS regressions will not work with censored data. Instrumental variables (IV) regression, which can also be used to deal with serial correlation, would be very complicated for the tobit regressions estimated here. As a rough check on the estimation, therefore, instrumental variable estimates to correct for serial correlation were performed only for the OLS regressions where the residuals were autocorrelated (namely, R&D in oil and gas recovery, coal gasification / liquefaction, and tar sands). The results of these regressions differ relatively little from those reported in Table 8. The coefficient on the lagged dependent variable is significant for each of the three types of R&D; and as in Table 8, the coefficient estimate is highest for oil and gas recovery R&D and lowest for tar sands R&D. (Additional information is available on request from the author).

The regression results presented thus far provide support for Hypotheses 1 and 4 advanced in this paper. The estimated coefficients on the lagged R&D intensity variables show that firms' current and past efforts in each R&D activity are positively correlated (Hypothesis 1), and that path dependence is high in four out of five of the R&D activities. Thus the positive autocorrelations between past and present R&D intensity demonstrated by the correlation coefficients hold up in the regressions as well. Additionally, the regressions indicate that the relationship between individual firms' R&D expenditures in one R&D activity and expenditures for the remainder of their R&D differs by R&D activity. This finding provides support for Hypothesis 4 that firms do not necessarily devote the same R&D effort to all R&D activities. Similarly, the differences between the mean R&D intensities for the five R&D activities discussed earlier and shown in Table 6 suggest that levels of effort differ between the five R&D activities.

6.3.3. Regression Results—Pooled R&D Activities. To provide further evidence regarding persistent differences in firms' R&D activities, the individual regression equations for the five R&D activities are stacked and run together in a single regression.[16] Table 9 reports Tobit results for the pooled regression. (Results obtained using OLS and standard SURE (seemingly unrelated regression estimation), appropriate for uncensored data, are similar to the Tobit results and are available on request from the author. This study does not attempt SURE for Tobit regression, due to its complexity.)

The pooled regression yields coefficient estimates similar to those reported in the separate regressions. Several of the earlier coefficient estimates on the lagged dependent variables had similar magnitudes; tests of equality between the various lagged R&D intensity variables in the pooled regression show that it is not possible to reject the hypothesis (at the 0.10 level using a two-tailed test) that the coefficients on lagged R&D intensity for oil and gas recovery, oil refining, and coal gasification / liquefaction are equal. Additional tests reject the hypothesis that the coefficients on all five lagged R&D intensity variables equal one another, and that the coefficients on lagged oil shale and tar sands R&D intensity equal one another. Other tests also reject equality between either oil shale or tar sands lagged R&D intensity and the other three lagged R&D intensity variables. (All tests are significant at the 0.05 level or less.)

The results from the individual R&D activity regressions also suggested that R&D intensity in a given R&D

[16] The following procedure is used to pool the five R&D activities into one regression. The original regression for each R&D activity is (omitting the year dummy variables and suppressing time and firm subscripts):

$$\text{R\&D}_i = a + b_i(\text{Lagged R\&D}_i) + c_i(\text{Lagged Asset}_i) + d_i(\text{Invest}_i) + f_i(\text{Total R\&D}_i) + g_i(\text{Rivalry}_i) + e_i$$

where $_i$ = type of R&D activity. Let $n = 1$ through N be the number of observations for the first original R&D activity regression, $N + 1$ through $2N$ the number of observations for the second regression, and so on. In the pooled regression, the dependent variable and right-hand side variables are: R&D = {R&D$_i$ if $n = 1, \ldots, N$; R&D$_j$ if $n = N + 1, \ldots, 2N$; R&D$_k$ if $n = 2N + 1, \ldots, 3N$; R&D$_l$ if $n = 3N + 1, \ldots, 4N$; R&D$_m$ if $n = 4N + 1, \ldots, 5N$; where $_i \neq j \neq k \neq l \neq$ m denote the five R&D activities}; Lagged R&D$_i$ = {Lagged R&D$_i$ if $n = 1, \ldots, N$; O otherwise}; Lagged R&D$_j$ = {Lagged R&D$_j$ if $n = N + 1, \ldots, 2N$; O otherwise}; and similarly for the other lagged R&D variables. The remaining right-hand side variables are treated in a similar fashion to the lagged R&D variables.

Table 9 Pooled R&D Activities—Tobit Regression[a] (t-statistics in parentheses)

Dependent Variable: R&D Activity $/Total Firm Sales $

Independent Variables					
Constant Term[b]		Oil Shale Invest. $/Sales $	0.0119† (0.6230)	Total Firm R&D Minus Coal Gas R&D $/Sales $	0.0259** (2.0594)
Lagged Oil & Gas R&D Intensity	1.0894* (33.4954)	Tar Sands Invest. $/Sales $	0.0063† (0.9695)	Total Firm R&D Minus Oil Shale R&D $/Sales $	0.0463* (3.7939)
Lagged Refining R&D Intensity	0.9911* (17.9298)	Lagged Petroleum Prod. Asset $/ Sales $	0.0002***† (1.7028)	Total Firm R&D Minus Tar Sands R&D $/Sales $	0.0412* (3.3367)
Lagged Coal Gasification/ Liquefaction R&D Intensity	1.0847* (12.1027)	Lagged Ref. Asset $/Sales $	0.0005† (0.9569)	Index of Rivalry in Oil & Gas R&D	−0.0018* (−2.7847)
Lagged Oil Shale R&D Intensity	0.7867* (28.7452)	Lagged Coal Mining and Gas./Lique. Assets $/Sales $	−0.0022***† (−1.6582)	Index of Rivalry in Refining R&D	−0.0021* (−3.4749)
Lagged Tar Sands R&D Intensity	0.4332* (2.9899)	Lagged Oil Shale Asset $/Sales $	0.0003† (0.0219)	Index of Rivalry in Coal Gas R&D	−0.0005** (−2.3082)
Petroleum Production Invest. $/Sales $	−0.0009***† (−1.6617)	Lagged Tar Sands Asset $/Sales $	0.0047*† (2.4657)	Index of Rivalry in Oil Shale R&D	−0.0018* (−3.4138)
Refining Invest. $/Sales $	0.0026† (1.1693)	Total Firm R&D Minus Oil & Gas R&D $/Sales $	0.0283** (2.2961)	Index of Rivalry in Tar Sands R&D	−0.0016* (−4.3817)
Coal Mining & Gas./Lique. Invest. $/Sales $	−0.0007† (−0.3289)	Total Firm R&D Minus Refining R&D $/Sales $	0.0424* (2.7753)	Year Dummy Variables	*
				χ^2 Statistic	3687.4892*
			N = 858		

[a] Regression includes data for the years 1975–81, except for refining R&D and coal gasification/liquefaction R&D which have data only for the years 1976–81 (due to missing data).

[b] Estimates of the constant term and of the year dummy variable coefficients are not reported separately. The significance level indicated for the year dummy variables refers to their joint significance in the regression.

* Significant at the 0.01 level or less.

** Significant at the 0.05 level or less.

*** Significant at the 0.10 level or less.

† Investment variables jointly significant at the 0.06 level. Lagged asset variables jointly significant at the 0.01 level. Investment and lagged asset variables together jointly significant at the 0.01 level.

Note: Tests for joint significance of investment and asset variables are χ^2 tests and significance levels are based on one-tailed tests. The significance level of the χ^2 statistic for the overall regression also is based on a one-tailed test. All other significance levels are based on two-tailed tests.

activity did not correlate strongly with investment and assets related to each R&D activity. The pooled regression results also support this conclusion. Although the investment variables related to all five R&D activities are jointly significant, as are all five asset variables (and as are all investment and asset variables together), the coefficient estimates on these variables are low, similar in magnitude to those in the individual R&D activity regressions. (Presumably standardized coefficients also would be similar in magnitude to the ones presented earlier.) As before, the lagged dependent variables may pick up much of any effect of the lagged asset variables. Evolutionary theory does not preclude ties to firms' heterogenous asset bases, and especially to the cumulative knowledge associated with them, as sources of persistent firm differences in R&D activity. The incorporation in the regression of the lagged dependent and lagged asset variables together, however, makes it difficult to ascertain the relative importance of these variables in generating persistent differences in R&D activities between and within firms.

6.3.4. Fixed Effect Regressions. The analysis presented thus far shows persistence in R&D activities through time as well as differences in the amount of effort devoted to the five R&D activities for the time period as a whole. The finding of both persistence and differences in R&D activity raises the possibility that if between and within firm differences in R&D activity are stable through time, this could produce the empirical finding of persistence in R&D activity. Hypothesis 2 deals with this issue. The earlier discussion regarding Hypothesis 2 suggested two possible sources of persistence in R&D: unchanging R&D activity through time or incremental changes in R&D activity. Hypothesis 2 posits the latter.

To test Hypothesis 2, this study uses what is known as "fixed effects" regression. This regression technique controls for that portion of the dependent variable, annual R&D intensity in this case, that remains fixed through time. If R&D intensity has a fixed portion, the incorporation of "fixed effects" in the regression will pick up some of the path dependence in R&D, causing the estimated coefficient on the lagged dependent variable (transformed as described below) to fall relative to that in the regression without fixed effects. The stronger the fixed effects, the greater the likely reduction in the coefficient.[17]

Three standard techniques can be used to estimate fixed effects regression. The first approach involves controlling for the fixed effects by adding time invariant dummy variables for each firm as independent variables in the regression. The other two approaches involve deleting the fixed effects from the regression, leaving only the time varying portion of the dependent and independent variables. Either of two methods will accomplish this: the use in the regression of first differences, or the use of the differences between each variable and its average per firm over the entire time period in the sample.[18] The latter two methods cannot be used for tobit regressions, however.[19]

All of these methods, however, suffer from the problem that in panel data with fixed effects, OLS and Tobit regressions produce inconsistent coefficients on the lagged dependent variable (Nickell 1981, Honore 1993).[20] Honore (1993) has developed a Tobit estimator for this situation that is consistent as the number of firms goes to infinity; unfortunately, Honore's estimator will not produce consistent results for the data in this sample, due to the relatively small number of firms.[21] The usual solution to the problem in OLS relies on instrumental variable estimation (see, e.g., Jacobson 1990). Tobit instrumental variables regression involving lagged dependent variables and fixed effects, however, is difficult to estimate.[22] In light of the difficulty of obtaining consistent Tobit estimates with fixed effects, the following analysis concentrates on OLS instrumental variables regression to incorporate fixed effects. The previous regressions without fixed effects showed that

[17] Fixed effects regressions also control for any unspecified firm-level effects on R&D expenditures. If these unspecified effects are correlated with any of the regressors, then controlling for fixed effects will eliminate bias in the regression coefficients (Hausman and Taylor 1981). Thus, the regression coefficients could also change from those in the non-fixed effects regressions due to elimination of bias in the estimated coefficients.

[18] Both of these techniques delete all constant terms from the regression, including firm-specific ones (i.e., the fixed effects).

[19] If a firm, for example, has zero expenditures in some periods and positive expenditures in others, the deviation from the time period average will be negative during the zero expenditure years. First differences also may produce negative values in some years. Tobit regression cannot handle negative values of the dependent variable.

[20] OLS produces a coefficient on the lagged dependent variable that has a downward, and often large, bias. If the other right-hand side variables are positively correlated with the lagged dependent variable, the coefficients on the other right-hand side variables have an upward bias (and vice versa when the variables are negatively correlated) (Nickell 1981). The direction of the tobit bias is not known. Although the general extent of the bias is not clear, it can be very large in some instances (Honore 1993).

[21] This is based on simulations performed by Honore for the author. Honore's technique also requires no serial correlation of the residuals, which may not apply here.

[22] This would also be difficult and time consuming to do using SAS, the only software available for use with these data.

the OLS coefficients were fairly close to the Tobit estimates.

To estimate the (OLS) fixed effects regressions, first differences were computed for all of the variables.[23] This procedure, however, creates a correlation between the first differenced lagged dependent variable and the error term.[24] To correct this general situation of correlation between a regressor and the error term, it is common to use an instrument for the regressor that is correlated with the regressor and but not with the error term. An obvious choice of an instrument for the first differenced lagged dependent variable is the dependent variable lagged two years (Anderson and Hsiao 1982, Jacobson 1990). Therefore, the dependent variable lagged two years was used as an instrument in the fixed effects OLS regressions.

For three types of the R&D activities, however, the two year lagged dependent variable may not be an appropriate instrument. Analysis of the non-fixed effects OLS regressions provided evidence discussed earlier of serial correlation in the residuals for oil and gas recovery, coal gasification / liquefaction, and tar sands R&D. The serial correlation raises the possibility that, for these types of R&D, the two year lagged dependent variable could be correlated with the error term in the fixed effects regression.[25] Therefore, a second set of fixed effects regressions uses the one and two year lagged values of the total firm R&D minus type of R&D activity variable as instruments for the first differenced lagged dependent variable, since these lagged "total" R&D variables are uncorrelated with the error term.[26]

Table 10 reports the results of the fixed effects regressions. Two stage least squares was used to obtain the instrumental variable estimates. Of greatest interest for Hypothesis 2 regarding incremental change (vs. fixed effects) are the coefficient estimates on the first differenced lagged dependent variables. Recall that if fixed effects are important, the coefficient estimate on the first differenced lagged dependent variable for each R&D activity should fall relative to the coefficient on the lagged dependent variable in the non-fixed effect regression. The direction of the results differ by type of R&D activity. When R&D intensity lagged two years is used as an instrument, the regressions produce the following results: for each of oil and gas recovery and refining R&D, the coefficient on the first differenced lagged R&D intensity variable is much lower than the coefficient on the lagged dependent variable in the initial OLS regression for each R&D activity; for each of coal gasification / liquefaction and tar sands R&D, the coefficient on the first differenced lagged dependent variable in the fixed effect regressions is greater than or equal to that on the lagged dependent variable in the initial OLS regression for each R&D activity; and for oil shale R&D, the coefficient estimate in the fixed effect regression is only slightly lower than in the initial OLS regression. All of the coefficient estimates on the first differenced lagged dependent variables in the fixed effects regressions are statistically significant at the 0.01 level or less except in the refining regression.

[23] In models with a lagged dependent variable, it is not appropriate to use dummy variables for individual firms to estimate fixed effects regression (Holtz-Eakin et al., 1988). The use of deviations from firm-level means is problematic, because instruments such as the dependent variable lagged more than one year will be correlated with the error term (also measured as the deviation from a firm-level mean). The use of first differences does create serial correlation of the errors, but because the use of instrumental variables produces consistent estimates (on uncensored data of course) as either the number of firms or time periods goes to infinity, first differencing is the preferred technique in this situation.

[24] To see this, suppose that the lagged dependent variable is the only regressor besides constant terms. Then first differencing yields the following regression equations:

$$(y_{i,t} - y_{i,t-1}) = b(y_{i,t-1} - y_{i,t-2}) + (e_{i,t} - e_{i,t-1})$$

where y = dependent variable, i = index denoting each firm, t = time period, and e = error term from the original (not differenced) regression. Note that $y_{i,t-1}$ and $e_{i,t-1}$ are correlated.

[25] Suppose the error term in the non-fixed effects regression is serially correlated, so that e_{t-1} is correlated with e_{t-2} in that regression. Then the first difference of the error term ($e_t - e_{t-1}$) is correlated with e_{t-2}. Since y_{t-2} (the dependent variable lagged two years) is by definition correlated with e_{t-2}, and if e_{t-1} and e_{t-2} are correlated, then y_{t-2} and e_{t-1} also may be correlated.

[26] By definition, these lagged "total" R&D variables are uncorrelated with either of the components of the estimated first differenced error term (e_t and e_{t-1}), since the residuals of an OLS regression are uncorrelated with the regressors. Other possible instruments, such as longer-term lags of the dependent variable, would have the disadvantage of reducing the number of observations in the IV regressions.

Table 10 OLS Fixed Effects Regressions[a] (t-statistics in parentheses)

Dependent Variable: First Differenced Oil & Gas Recovery R&D/Total Firm Sales $
Instrument(s) for Lagged Dependent Variable: Regression No. 1: Oil & Gas Recovery R&D $/Sales $, Lagged 2 years
Regression No. 2: Total Firm R&D Minus Oil & Gas R&D/Sales $, Lagged 1 year and Lagged 2 years

Regression Number[b]	Lagged Dependent Variable	First Differenced Petroleum Production Invest. $/Sales $	First Differenced Lagged Petroleum Prod. Assets $/Sales $	First Differenced Total Firm R&D Minus Oil & Gas R&D $/Sales $	First Differenced Index of Rivalry in Oil & Gas R&D	R^2	N
1	−0.8411***	0.0004	0.0001	0.0376	−0.0143**	0.1768	156
	(−1.795)	(0.915)	(0.255)	(1.154)	(−2.096)		
2	−3.3176	0.0013	0.0006	0.0522	0.0010	0.0324	156
	(−1.174)	(0.875)	(0.430)	(0.560)	(0.039)		

Dependent Variable: First Differenced Refining R&D $/Total Firm Sales $
Instrument(s) for Lagged Dependent Variable: Regression No. 1: Refining R&D $/Sales $, Lagged 2 years
Regression No. 2: Total Firm R&D Minus Refining R&D $/Sales $, Lagged 1 year and Lagged 2 years

Regression Number[b]	Lagged Dependent Variable	First Differenced Refining Invest. $/Sales $	First Differenced Lagged Ref. Assets $/Sales $	First Differenced Total Firm R&D Minus Ref. R&D $/Sales $	First Differenced Index of Rivalry in Refining R&D	R^2	N
1	0.1422	−0.0005	−0.0020	0.0173	−0.0086*	0.1198	130
	(0.414)	(−0.305)	(−1.228)	(0.637)	(−3.221)		
2	−3.4500	0.0008	0.0058	−0.0113	−0.0106	0.0168	130
	(−0.910)	(0.129)	(0.598)	(−0.111)	(−1.072)		

Dependent Variable: First Differenced Coal/Gas Liquef. R&D/Total Firm Sales $
Instrument(s) for Lagged Dependent Variable: Regression No. 1: Coal Gasification/Liquefaction R&D $/Sales $, Lagged 2 years
Regression No. 2: Total Firm R&D Minus Coal Gas./Liquef. R&D $/Sales $, Lagged 1 year and Lagged 2 years

Regression Number[b]	Lagged Dependent Variable	First Differenced Coal Mining and Gas./Liquef. Invest. $/Sales $	First Differenced Lagged Coal Mining and Gas./ Liquef. Assets $/Sales $	First Differenced Total Firm R&D Minus Coal Gas. R&D $/Sales $	First Differenced Index of Rivalry in Coal Gas R&D	R^2	N
1	1.0488***	0.0008	0.0006	0.0295	−0.0024**	0.1378	130
	(1.900)	(0.624)	(0.353)	(0.787)	(−2.380)		
2	1.1810	0.0007	−0.00005	0.0214	−0.0019	0.0736	130
	(1.221)	(0.429)	(−0.017)	(0.396)	(−1.176)		

Dependent Variable: First Differenced Oil Shale R&D/Total Firm Sales $
Instrument(s) for Lagged Dependent Variable: Regression No. 1: Oil Shale R&D $/Sales $, Lagged 2 years
Regression No. 2: Total Firm R&D Minus Oil Shale R&D/Sales $, Lagged 1 year and Lagged 2 years

Regression Number[b]	Lagged Dependent Variable	First Differenced Oil Shale Invest. $/Sales $	First Differenced Lagged Oil Shale Assets $/Sales $	First Differenced Total Firm R&D Minus Oil Shale R&D $/Sales $	First Differenced Index of Rivalry in Oil Shale R&D	R^2	N
1	0.6070*	0.0091	−0.0786	0.0187	−0.0160*	0.4162	156
	(5.225)	(0.403)	(−2.230)	(0.435)	(−6.940)		
2	1.1144***	0.0123	−0.1069***	0.0396	−0.0145*	0.2264	156
	(1.735)	(0.390)	(−1.781)	(0.611)	(−3.892)		

Dependent Variable: First Differenced Tar Sands R&D $/Total Firm Sales $
Instrument(s) for Lagged Dependent Variable: Regression No. 1: Tar Sands R&D $/Sales $, Lagged 2 years
Regression No. 2: Total Firm R&D Minus Tar Sands R&D/Sales $, Lagged 1 year and Lagged 2 years

Regression Number[b]	Lagged Dependent Variable	First Differenced Tar Sands Invest. $/Sales $	First Differenced Lagged Tar Sands Assets $/Sales $	First Differenced Total Firm R&D Minus Tar Sands R&D $/Sales $	First Differenced Index of Rivalry in Tar Sands R&D	R^2	N
1	0.2103	−0.0008	−0.0025	−0.0009	−0.0015*	0.1571	156
	(2.130)	(−0.225)	(−0.870)	(−0.055)	(−3.874)		
2	0.3526	−0.0008	−0.0027	−0.0004	−0.0014*	0.1220	156
	(0.756)	(−0.214)	(−0.860)	(0.025)	(−3.358)		

[a] Instrumental variables regression estimated via two stage least squares. R^2 reported from the second stage regression.

[b] Both regressions also include first differenced year dummy variables as controls. Since few of the coefficients on the year dummy variables are significant at the 0.05 level or less, they are not reported in the table.

* Significant at the 0.01 level or less.

** Significant at the 0.05 level or less.

*** Significant at the 0.10 level or less.

This first set of fixed effects regressions suggests that R&D intensity in coal gasification / liquefaction, tar sands and oil shale primarily reflects incremental change. For each of these three R&D activities, the coefficient estimate on the first differenced lagged dependent variable either falls very little or rises relative to the coefficient on the lagged dependent variable in the non-fixed effect regression. In the refining R&D intensity fixed effects regression, the estimated coefficient on the first differenced lagged dependent variable drops to 0.14 and is not significantly different from zero, consistent with a strong fixed effect. Finally, the negative coefficient of -0.84 in the oil and gas recovery R&D equation suggests that once the fixed effect is removed from the equation (by using first differences), R&D intensity cycles up and down from year to year. This resulting converging cyclicality (rather than a trend) suggests that fixed effects have a strong impact on oil and gas recovery R&D intensity.

The second set of fixed effects regressions, using one and two year lagged values of the total R&D minus type of R&D activity variables as instruments, shows a similar pattern. For each R&D activity except refining, the coefficient estimate on first differenced lagged R&D intensity has the same sign and is greater in absolute value than in the first fixed effects regression; the coefficient estimate for refining turns negative. None of the coefficient estimates for first differenced lagged R&D intensity are significant except in the oil shale regression.[27] The result for oil shale is strengthened, since the coefficient on first differenced lagged R&D intensity is significant and rises above that for lagged R&D intensity in the initial OLS regression.

The initial set of fixed effects regressions suggests greater fixed effects in oil and gas recovery and refining R&D expenditures than in the other R&D activities, where the results point to incremental change. Although these results make it tempting to conclude that the R&D activities having greater observed path dependence also have greater fixed effects, the regression results do not fully support for such a conclusion. In particular, the initial OLS regressions indicate that both coal gasification / liquefaction and oil shale R&D have a relatively high degree of path dependence, yet the fixed effect regressions support the hypothesis of incremental change. Overall, the fixed effects regressions provide some support for the incremental adaptation view of evolutionary theory and some support for the greater degree of fixity hypothesized in organizational ecology. The mixed results point to the need for additional study of the speed of change of R&D and other activities within firms.

7. Summary and Conclusion

The analysis of R&D expenditures in the petroleum industry provides a good deal of support for the four hypotheses generated from evolutionary theory. A comparison of correlation coefficients for R&D activity expenditures with correlation coefficients for related types of investment spending shows much greater path dependence in R&D, providing support for Hypothesis 1 that R&D tends to be path dependent. In the regressions, the positive and significant coefficient estimates on lagged R&D intensity add further support for Hypothesis 1. The evidence on differences in levels of effort across R&D activities between and within firms also supports evolutionary theory. In each R&D activity, the relatively high coefficient of variation for the firm-level average R&D intensity, as compared with the lower coefficient of variation for the industry-level average R&D intensity, provides support for Hypothesis 3 that firms differ in their level of effort in an individual R&D activity. The differences between the mean firm-level average R&D intensities of the five R&D activities support Hypothesis 4 that each individual firm devotes dissimilar levels of effort to the various R&D activities; differences across R&D activities in the coefficients of variation for the firm-level average R&D "share" variables supports the additional contention in Hypothesis 4 that the pattern of within firm differences in R&D activity also differs between firms. The regression results further buttress the findings regarding Hypothesis 4, showing that the relationship between the R&D inten-

[27] This may arise due to loss in efficiency because the "total" R&D variables used as instruments are less highly correlated with lagged R&D intensity (a component of the first differenced lagged R&D intensity variable requiring an instrument) in each R&D activity than is two-year lagged R&D intensity.

sity of an activity and the R&D intensity of the remainder of the firm's R&D differs across R&D activities. With regard to hypothesis 2 on the degree of incremental change in R&D efforts, the regression analysis provides evidence of incremental adaptation in some R&D activities and greater fixity in other activities.

Hypotheses 1, 3, and 4, *in combination*, provide a test of the proposition derived from evolutionary theory that firms within an industry exhibit persistent differences in their R&D activities. In addition, Hypothesis 2 suggests that persistence in R&D activities reflects incremental adaptation and change, as posited by evolutionary theory. Although some modifications to orthodox economic models might be used to explain each one of the hypotheses individually (e.g., adaptive expectations, adjustment costs, or heterogeneous firm assets), evolutionary theory provides a single, consistent explanation for all four hypotheses. Thus, it is important to examine the empirical results of this study as a whole, rather than as tests of separate hypotheses. Taken together, the empirical evidence points to *persistent differences* in petroleum firms' R&D activities, and to incremental adaptation and change in some of these activities.

The results of this study, while supportive of persistent differences in firm activities, constitute but a first step in the empirical study of the topic. Many issues remain to be explored, with regard to R&D and to other firm activities. For example, the sources of initial heterogeneity in firms' R&D or other knowledge bases bear examination, as do the way in which the initial knowledge bases change in reaction to changes in the external environment (e.g., price changes). The latter requires an empirical evolutionary model of firms' interactions with and responses to the external selection environment, a complex undertaking not attempted here. Another extension of this study would involve further analysis of the rate of incremental adaptation and change in R&D as well as in other activities within the firm. Additionally, it would be helpful to analyze the joint evolution of bundles of activities within the firm. For example, this study tested for the importance to R&D activity of ties to related firm-level assets and investment. While the investment and asset variables were jointly significant, in most of the regressions the variables had relatively low coefficient estimates (even when standardized). Path dependence in R&D does not preclude ties of R&D to related firm resources, or the joint evolution of R&D and other firm resources. As a first step, the analysis performed here could incorporate longer term lags of both R&D expenditures and of related investment and assets to examine the structure of the persistence for all three variables over time. It might also help to separate the analysis of investment and assets from that of lagged R&D.

In summary, the *persistent differences* in petroleum firms' R&D activities documented here suggest that the associated knowledge bases change only slowly. These findings may generalize to other types of organizational knowledge as well. In particular, evolutionary theory may help to provide an explanation for persistent heterogeneity of firms' know-how and capabilities, which in turn may produce differential firm strengths (or weaknesses) that change only slowly over time. These slow-to-change differential firm capabilities can produce persistent differences in firms' performance.[28]

[28] I am indebted to Arthur Andersen of the U.S. Department of Energy (DOE) for allowing me access to the Financial Reporting System database, and to Jon Rasmussen and Crawford Honeycutt, also of DOE, for invaluable help and advice. I also benefitted from the econometric advice of Joe Altonji, Bo Honore, Bruce Meyer, and Dan Sullivan, and the research assistance of Sonal Desai and Ron Konezny. Bruce Kogut, Marvin Lieberman, Marshall Meyer, Cynthia Montgomery, Richard Nelson, Sidney Winter, and seminar participants at the University of Pennsylvania and New York University provided helpful comments. In addition, two anonymous reviewers and an associate editor provided extremely helpful feedback. The Reginald H. Jones Center of the Wharton School, University of Pennsylvania, and the J. L. Kellogg Graduate School of Management at Northwestern University, provided travel funds for this research. Any errors are mine alone.

References

Amit, Rafael and Paul Schoemaker, "Strategic Assets and Organizational Rent," *Strategic Management J.*, 14 (1992), 33–46.

Anderson, T. W. and Cheng Hsiao, "Formulation and Estimation of Dynamic Models Using Panel Data," *J. Econometrics*, 18 (1982), 47–82.

Barney, Jay B., "Firm Resources and Sustained Competitive Advantage," *J. Management*, 17 (1991), 99–120.

Cohen, Wesley M. and Daniel A. Levinthal, "Absorptive Capacity:

A New Perspective on Learning and Innovation," *Administrative Sci. Quarterly*, 35 (1990), 128–152.

Conner, Kathleen R., "A Historical Comparison of Resource-Based Theory and Five Schools of Thought Within Industrial Organization Economics: Do We Have a New Theory of the Firm?" *J. Management*, 17 (1991), 121–154.

Cyert, Richard M. and James G. March, *A Behavioral Theory of the Firm*, (2nd Ed.), Blackwell, Cambridge, MA, 1992.

David, Paul A., "Path Dependence: Putting the Past Into the Future of Economics," Working Paper, Stanford University, Stanford, CA, 1988.

Dosi, Giovanni, "Sources, Procedures, and Microeconomic Effects of Innovation," *J. Economic Literature*, 26 (1988), 1120–1171.

——, David Teece, and Sidney Winter, "Toward A Theory of Corporate Coherence: Preliminary Remarks," in Giovanni Dosi, Renato Giannetti, and Pier Angelo Toninelli, (Eds.), *Technology and Enterprise in Historical Perspective*, Clarendon Press of Oxford University Press, New York, 1992, 185–211.

Fair, Ray C., "A Note on the Computation of the Tobit Estimator," *Econometrica*, 45 (1977), 1723–1727.

Freeman, Christopher, *The Economics of Industrial Innovation*, 2nd edition, MIT Press, Cambridge, MA, 1982.

Grabowski, Henry G., "The Determinants of Industrial Research and Development: A Study of the Chemical, Drug, and Petroleum Industries," *J. Political Economy*, 60 (1968), 292–306.

Hall, Bronwyn and Fumio Hayashi, "Research and Development as an Investment," NBER Working Paper #2973, National Bureau of Economic Research, Cambridge, MA, 1989.

Hannan, Michael T. and John Freeman, *Organizational Ecology*, Harvard University Press, Cambridge, MA, 1989.

Hausman, Jerry A. and William E. Taylor, "Panel Data and Unobserved Individual Effects," *Econometrica*, 49 (1981), 1377–1398.

Helfat, Constance E., "Firm-Specificity in Corporate Applied R&D," *Organization Sci.*, 5 (1994), 173–184.

Himmelberg, Charles and Bruce Peterson, "R&D Investment and Internal Finance: A Panel Study of Small Firms in High-Tech Industries," Working Paper, affil., city, state, 1991.

Holtz-Eakin, Douglas, Whitney Newey, and Harvey S. Rosen, "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56 (1988), 1371–1395.

Honore, Bo, "Orthogonality Conditions for Tobit Models with Fixed Effects and Lagged Dependent Variables," *J. Econometrics*, 59 (1993), 35–61.

Jacobson, Robert, "Unobservable Effects and Business Performance: Reply to the Comments of Boulding and Buzzell," *Marketing Sci.*, 9 (1990), 92–95.

Kogut, Bruce, "Joint Ventures: Theoretical and Empirical Perspectives," *Strategic Management J.*, 9 (1988), 319–332.

Levin, Richard C., Alvin K. Klevorick, Richard K. Nelson, and Sidney G. Winter, "Appropriating the Returns from Industrial Research and Development," *Brookings Papers on Economic Activity*, (1988), 783–831.

Lippman, S. A. and R. P. Rumelt, "Uncertain Imitability: An Analysis of Interfirm Differences in Efficiency Under Competition," *Bell J. of Economics*, 13 (1982), 418–438.

Maddala, G. S., *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge, England, 1983.

Mahoney, Joseph T. and J. Rajendran Pandian, "The Resource-Based View Within the Conversation of Strategic Management," *Strategic Management J.*, 13 (1992), 363–380.

Meyer, Marshall W., "Notes of a Skeptic: From Organizational Ecology to Organizational Evolution." In Jitendra V. Singh (Ed.), *Organizational Evolution: New Directions*, Sage Publications, Newbury Park, CA, 1990, 298–314.

Myers, S. and N. S. Majluf, "Stock Issues and Investment Policy When Firms Have Information That Investors Do Not Have," *J. Financial Economics*, 13 (1984), 187–222.

Nadiri, M. Ishaq and George C. Bitros, "Research and Development Expenditures and Labor Productivity at the Firm Level: A Dynamic Model," in John W. Kendrick and Beatrice N. Vaccara, (Eds.), *New Developments in Productivity Measurement and Analysis*, University of Chicago Press, Chicago, 1980, 387–417.

Nelson, Richard R., "Uncertainty, Learning, and the Economics of Parallel Research and Development Efforts," *Review of Economics and Statistics*, 43 (1961), 351–364.

——, "Capitalism As an Engine of Progress," *Research Policy*, 19 (1990), 192–214.

——, "Why Do Firms Differ and How Does It Matter?" *Strategic Management J.*, Winter Special Issue (1991), 61–74.

—— and Sidney G. Winter, *An Evolutionary Theory of Economic Change*, The Belknap Press of Harvard University Press, Cambridge, MA, 1982.

Nickell, Stephen, "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49 (1981), 1417–1426.

Peteraf, Margaret A., "The Cornerstones of Competitive Advantage: A Resource-Based View," *Strategic Management J.*, 14 (1993), 179–191.

Polanyi, Michael, *The Tacit Dimension*, Doubleday Anchor, Garden City, NY, 1967.

Rasmussen, Jon A., "Applications of a Model of Endogenous Technical Change to U.S. Industry Data," *The Review of Economic Studies*, 35 (1973), 225–238.

Rumelt, Richard P., "Towards a Strategic Theory of the Firm," in B. Lamb (Ed.), *Competitive Strategic Management*, Prentice-Hall, Englewood Cliffs, NJ, 1984, 556–570.

Schankerman, Mark and M. Ishaq Nadiri, "Investment in R&D, Costs of Adjustment, and Expectations," in Zvi Griliches, (Ed.), *R&D, Patents, and Productivity*, University of Chicago Press, Chicago, 1984, 315–338.

Scherer, F. M., "Schumpeter and Plausible Capitalism," *J. Economic Literature* 30 (1992), 1416–1433.

Scott, W. Richard, *Organizations: Rational, Natural, and Open Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1987.

Simon, Herbert A., "Rationality as Process and Product of Thought," *American Economic Review*, 68 (1978), 1–15.

Singh, Jitendra V., "Future Directions in Organizational Evolution," in Jitendra V. Singh (Ed.), *Organizational Evolution: New Directions*. Sage Publications, Newbury Park, CA, 1990, 315–319.

Stiglitz, Joseph E. and Andrew Weiss, "Credit Rationing in Markets with Imperfect Information," *American Economic Review*, 71 (1981), 393–410.

Teece, David J., Gary Pisano, and Amy Shuen, "Dynamic Capabilities and Strategic Management," Working Paper, University of California at Berkeley, 1992.

Theil, Henri, *Principles of Econometrics*, John Wiley and Sons, New York, 1971, 556–562.

U.S. Department of Energy, *Energy Company Development Patterns in the Postembargo Era*, Vol. 1, U.S. Government Printing Office, Washington, D.C., 1982.

Wernerfelt, Birger, "A Resource-Based View of the Firm," *Strategic Management Journal*, 5 (1984), 171–180.

Winter, Sidney G., "Knowledge and Competence as Strategic Assets," in David J. Teece (Ed.), *The Competitive Challenge: Strategies for Industrial Innovation and Renewal*, Ballinger, Cambridge, MA, 1987.

——, "Survival, Selection, and Inheritance in Evolutionary Theories of Organization," in Jitendra V. Singh (Ed.), *Organizational Evolution: New Directions*, Sage Publications, Newbury Park, CA, 1990.

Accepted by Richard M. Burton; received May 18, 1992. *This paper has been with the authors* 6 *months for* 2 *revisions*.