INVESTMENT IN ENERGY EFFICIENCY: DO THE CHARACTERISTICS OF FIRMS MATTER?

by

Stephen J. DeCanio

and

William E. Watkins

Department of Economics
University of California
Santa Barbara, CA 93106
Voice: 805-893-3130
Fax: 805-893-8830
e-mail: decanio@econ.ucsb.edu
bill@econ.ucsb.edu

October 17, 1995; Revised April 26, 1996; Final Revision July 13, 1996

This research was supported in part by a grant from the U.S. Environmental Protection Agency. Capable research assistance was provided by Paymon Aliabadi. We appreciate the helpful comments of Roderick Little, Ron Schoenberg, Seema Arora, Douglas Steigerwald, and Perry Shapiro. We appreciate the constructive comments of three anonymous referees. The views expressed are those of the authors alone, who are also responsible for any errors.

ABSTRACT

The literature on energy efficiency provides numerous examples of apparently profitable technologies that are not universally adopted. Yet according to the standard neoclassical theory of investment, profit-maximizing firms should undertake all investments with a positive net present value. The standard theory also holds that the discount rate for computing the present value of a project should be the return available on other projects in the same risk class, and therefore should not depend on characteristics of the firm. This model as applied to energy-saving investments is tested by examining whether firms' characteristics influence their decision to join the Environmental Protection Agency's voluntary Green Lights program. A discrete choice regression is estimated over a large sample of participating and non-participating firms. Missing values in the data matrix are replaced with multiple imputations from a distribution estimated using the EM algorithm. The results show that: (1) substantial improvements in the power of hypothesis tests can be achieved through maximum likelihood imputation of missing data, and (2) contrary to the conventional theory, the characteristics of firms do affect their decision to join Green Lights and commit to a program of investments in lighting efficiency.

Journal of Economic Literature Classification Numbers: C10, D21, G31, L20, Q40.

I. Introduction and Background---Implications of the Neoclassical Theory of Investment.

If the neoclassical theory of the firm were like the theory of the propagation of light through the ether, then the unwillingness of allegedly profit-maximizing firms to invest in energy efficiency would be comparable to the Michelson-Morley experiment. The slow diffusion of energy-efficiency technologies---in lighting, motors, and HVAC systems---is one of the most significant empirical challenges to the conventional view that rational firms make all investments with a positive net present value. A sizable technical and scholarly literature presents an imposing body of evidence that the typical rate of return available from energy-efficiency investments is much higher than the discount rate for projects of comparable risk.[1]

Nevertheless, economics is not physics, and the true significance of these findings is controversial. Economists in general are skeptical about the existence, let alone the ubiquity, of unrealized profit opportunities. The full range of energy investment costs includes difficult-to-measure components such as transactions costs, monitoring costs, administrative costs, and adjustment costs. Rules of thumb that seem to foreclose a range of profitable investments may be in place to deal with other, more intractable and more important, problems of control and incentive (Antle and Eppen 1985). Recent theoretical work has also stressed the option value of putting off investments until more information on prices and costs arrives, with the implication that the suitable discount rate for calculating NPV should be higher than had been thought (Bernanke 1983; Kester 1984; Pindyck 1991; Dixit 1992; Hassett and Metcalf 1993; Dixit and Pindyck 1994), but Sanstad, Blumstein, and Stoff (1995) have shown that this effect is much too small empirically to account for the apparent discrepancy between estimated implicit discount rates for energy-saving investments and the appropriate cost of capital comparison rate.

It is therefore not easy to contemplate designing a common numerical measure of the performance of investments in energy-efficiency technologies that would enable the conventional investment hypothesis to be tested directly. There is, however, another indicator that enables a very clean and decisive test to be carried out. The U.S. Environmental Protection Agency has recently begun to institute a series of voluntary pollution prevention programs. Firms that join these programs are not subjected to any differential regulation by the EPA; instead, the firms that join are able to take advantage of EPA's technical expertise and can enjoy the benefits of whatever favorable publicity comes from their participation.[2] Examples include Green Lights, the 33/50 Program, and the Energy Star programs. Some of these have been in operation for a sufficient length of time to generate a considerable amount of information on voluntary pollution prevention by firms. The decision to join Green Lights, for example, may be thought of as a signal (albeit a noisy one) of a firm's willingness to undertake energy conservation investments. According to the standard theory outlined above, *the decision to join Green Lights should not be associated systematically with characteristics of the firms that join.*

Thus, instead of trying to compare the level of the returns (which are subject to the problems of measurement and interpretation described above) to some suitable target rate, it ought to be possible to use the fact that *the investment decision rule should be invariant across firms* to devise an alternative and more powerful test of the conventional theory. One of the clearest consequences of the neoclassical theory of investment is that the discount rate used to calculate NPVs should be the return available on *projects* in the same risk class, which does not depend on the characteristics of the *firms* contemplating the investment. Because energy-efficiency investments are frequently generic (improving the lighting efficiency of office space, for example) the go/no-go decision on these investments should be independent of any other characteristics of the firms undertaking them.

It is worth taking a moment to see just how sharp this implication of the conventional theory is. The statistical test that will be performed does not in any way hinge on the undoubted facts that some energy-saving investments will be more profitable than others, or that some firms will have a greater selection of such profitable projects because, for example, they face higher electricity prices or they have been slow in beginning to implement the new technologies. Consider two multi-facility firms, A and B, with A being the firm facing higher electricity prices and having more opportunities for profitable energy savings. Rank the energy-saving investments available to each firm according to their internal rates of return,[3] and plot these rates against the cumulative dollars that would be required to make the investments. The relationships would be as shown in Figure 1.

{Insert Figure 1}

If r_0 is the rate of return available on projects of comparable risk, Company A would invest I_A, while Company B would invest I_B. However, the fact that Company A had either a greater number of or a more profitable set of projects does not obviate the fact that Company B should also, under the conventional NPV > 0 investment rule, be investing in energy efficiency. Both companies would be motivated to join the Green Lights program, regardless of their different circumstances. Even a company with *no* lighting investments meeting the profitability criterion[4] would have no reason not to join

Green Lights, because the program is entirely voluntary and provides publicity benefits and access to information and expertise.

II. The Green Lights Program and the Data.

Green Lights was launched in January, 1991. It is the flagship of the EPA's voluntary pollution prevention programs. The Global Change Division (now called the Atmospheric Pollution Prevention Division) of EPA targeted lighting systems first because (1) lighting is part of the energy budget of all private-sector and public enterprises, (2) lighting constitutes a significant segment of total electricity demand in the U.S. (20-25% according to *Building Operating Management* (1992)), and (3) EPA's analysis showed that very high rates of return were available by retrofitting with a simple set of proven technologies. These included electronic ballasts, energy-efficient fluorescent bulbs, occupancy sensors, rationalized maintenance and replacement schedules, and replacement of obsolete reflectors and lenses with more modern fixtures.

Firms can join Green Lights after being approached by EPA or on their own initiative. (The program has received considerable publicity since its inception.) Both publicly traded and privately held firms are members. Public sector and non-profit organizations, such as state and local governments, schools, and hospitals are also free to join. A company joins Green Lights by signing a Memorandum of Understanding (MOU) with the EPA. The firm commits to surveying its facilities and, within five years, making all lighting upgrades that meet a clear profitability test.[5] In return, EPA provides technical expertise, including a computerized expert system that can be used to analyze possible upgrade projects, information on the availability of rebates and sources of financing, a clearinghouse for information provided by other participating firms and by Green Lights “allies” (both utilities and sellers of modern lighting equipment), and publicity materials for participants that wish to receive it. The program is wholly voluntary, with no penalty for withdrawal at any time and no added regulatory burden for joining. A few firms have withdrawn or been dropped from the program. Written into the MOU is a provision that no upgrade need be undertaken that does not maintain or improve the working environment. Participants agree to minimal annual reporting, and to designation of a manager with responsibility to serve as the contact person between the EPA and the company. The Green Lights liaison (designated the “Green Lights Implementation Director” or GLID) ordinarily has multiple other responsibilities within the firm.

In terms of program growth, Green Lights appears to be successful. As of mid-1995, over 1700 companies had joined, representing 4.7 billion square feet of lighted space. The rate of growth in the number of participants has been 7% per annum, with no drop-off in the rate of growth visible in the data (U.S. EPA 1995). More to the point of the present paper, however, the decision to join Green Lights is a clear decision---either an MOU is signed or it is not---and this decision can be tested for association with other characteristics of the firm.

Our strategy is to identify the Green Lights partners in a large database containing firm-specific economic and financial information. We chose the Disclosure database of information provided to the SEC by publicly listed corporations (Disclosure, Inc. 1993). This contains information on over 9000 U.S. corporations, 268 of which had joined the Green Lights program as of the beginning of 1995. By performing a logistic regression (with dependent variable a categorical variable indicating whether or not a company had joined Green Lights), it is possible to test the hypothesis of the irrelevance of firm characteristics, and to explore the magnitude and direction of influence of a number of possible explanatory variables as well.

III. Statistical Issues.

1. The Problem of Missing Data.

The biggest statistical difficulty in carrying out the analysis is that the Disclosure data contains many observations with missing values for one or more of the variables of interest. Of the 9548 companies present in the database, a typical set of explanatory variables (see below) has only about 1000 "complete" observations, that is, observations in which all variables are non-missing.[6] On the other hand, only about one-third of the data points are actually missing, so that 2/3 of the potential total raw information is present. Obviously, the standard statistical procedure of carrying out the analysis only on those observations that are complete would throw away a very large portion of the information that could be brought to bear.

Using only complete observations carries inferential pitfalls as well. It is possible that the missing data are not missing completely at random; for example, suppose that smaller firms are less likely to be represented by complete sets of observations than large ones, but that the missing small firm observations are missing at random. This is Missing at Random (MAR), but not Missing Completely at Random (MCAR)[7] and can result in erroneous inferences. If size is an important explanatory variable (as it might be if the EPA's "marketing" effort was geared to approach larger firms first), the covariation between size and "missingness" could mislead the investigation if all observations with any missing values were simply dropped from the sample.

The question of how best to deal with missing data has a long history in the statistics literature, although it has received somewhat less treatment in economics. The first issue is whether the standard estimators provide consistent parameter estimates using only the complete observations. Griliches (1986) provides an answer that does not require that the missing observations be random. The missing observations are *ignorable* "as long as the conditional expectation of Y given X does not depend on which X's are missing" (p. 1487). That is, the missing observations are ignorable if including them would not change the expectations of the parameter estimates.[8] While the condition for ignorability is weaker than randomly missing observations and does allow systematically missing observations, no test for ignorability has been developed. The concepts of MCAR and MAR are more straightforward. MCAR means exactly that, the data are missing completely at random. MAR means that the missing data depend on the values of the other explanatory variables. Thus, all MCAR processes and a subset of MAR processes are ignorable. MCAR is ignorable and MAR may be.[9] In the estimation procedure that follows, we assume the missing data are ignorable for the theoretical reasons discussed below in Section 2c.

Rather than attempting to develop a test for ignorability, most of the literature takes the approach of devising methods by which the missing observations can be estimated conditional on the data being ignorable, MCAR, or MAR. Greene (1993), Little and Rubin (1987), and Griliches (1986) indicated that replacing missing observations in the data matrix X by consistently estimated observations would not bias an ultimate OLS model estimator. Of course, a consistent estimator for the missing x_i's is dependent on the relationship of the missing data and the remaining variables. Under different assumptions, replacement observations can be consistently estimated by OLS, GLS, or Maximum Likelihood.

These procedures can result in gains in the efficiency of the ultimate estimator of the parameters of the substantive model. The importance of these efficiency gains depends on the particular problem; in our case, it seems likely that the gains would be large because the missing observations are not confined to one variable, the fraction of complete observations relative to the entire data base is small, and the percentage of firms in the dataset that joined Green Lights is small. We have missing observations that are spread among 11 of the 26 independent variables. Only 1,033 of the 9,548 observations have complete data, and our ultimate estimators are LOGIT.[10] Only about one third of the data points are missing, but about 90 percent of the observations are affected.

2. *Methods.*

a. Potential Estimators of Missing Data.

Ignoring the missing data, substituting unconditional means, "hot deck" methods (described below), and estimation of the missing data by weighted or unweighted OLS or GLS have been shown by Little (1992), Rubin and Schenker (1986), Wang, Sedransk, and Jinn (1992), and others not to be appropriate methods of coping with missing data, except under very restrictive conditions. As mentioned previously, ignoring the data with missing observations is inefficient at best.

Substituting unconditional means for the missing values has the obvious drawback that the explanatory variables will show less variance than they should, but the problems with this technique are more serious than that. This method biases both the coefficient estimates and variances of those estimates when the missing *x*'s are not independent of the other independent variables.

Hot deck methods, including extensions such as Bayesian Bootstrap, are commonly used by government agencies for imputing missing data points in data sets to be used by other investigators (Little and Rubin 1987). In its simplest form, this method replaces missing data points by data drawn at random, with replacement, from the existing observations. Extensions include drawing with replacement from the estimated distribution of possible values, stratifying respondents[11] prior to imputing data with samples drawn with replacement from the same stratum, and drawing with replacement from an estimated distribution of the stratum's possible values. Little and Rubin (1987), Little (1992), and Rubin and Schenker (1986) provide an introduction to the literature on these methods. The more sophisticated of such methods have been shown to have some desirable properties, and they can be used as a basis for multiple imputation.[12] However, they are inferior to multiple imputation using draws from a distribution where the distribution parameters are estimated by a maximum likelihood estimator that takes account of interrelationships among the variables.

OLS, GLS, weighted OLS, or weighted GLS using the available information in the X matrix may provide consistent estimates of the missing data. However, consistency is dependent on restrictive assumptions about the missingness of the data (Little 1992), and calculation of the true variance of the model parameter estimates is difficult. We concluded from our review of the literature that Multiple Imputation of draws from a distribution parameterized by a Maximum Likelihood estimator provided the best solution to the problem of missing observations. We found that indeed the use of the full data set led to substantial improvements in efficiency (and more reliable inference) over using only the complete observations or a hot deck procedure.

b. Consistency of Binary Choice Models When Missing Values Are Imputed.

Before proceeding to a description of the method by which the missing values were estimated, it is necessary to point out that if the missing values are imputed properly, then Logit or Probit models have the same consistency properties as linear OLS with imputed data. Most of the results in the literature are derived for OLS models, but it is an easy generalization to show that methods using consistent estimators to impute the missing values will also yield consistent estimators of binary choice models. A binary choice model can be thought of as a nonlinear function of an OLS model.[13] The log likelihood function for the binary choice model is of the standard type:

$$Prob\ (Y = y_i) \quad = \quad F(\$x_i) \ \text{ if } \ y_i = 1, \ \text{ and} \qquad \text{(1a)}$$

$$Prob\ (Y = y_i) \quad = \quad [1 - F(\$x_i)], \text{ if } \ y_i = 0 \qquad \text{(1b)}$$

In this formulation, the likelihood and log likelihood functions are given by

$$L = \prod_i \, [F(\beta x_i)]^{y_i} \, [1 - F(\beta x_i)]^{1-y_i} \tag{2}$$

$$ln\, L = \sum_i \{[y_i \, ln\, F(\beta x_i)] + (1 - y_i)\, ln\, [1 - F(\beta x_i)]\} \tag{3}$$

Here F is the cumulative distribution function associated with either a logit or probit specification. The first-order condition for a maximum is:

$$\sum_i \{[y_i - F(bx_i)]\, f(bx_i)\, x_i \,/\, [F(bx_i)(1 - F(bx_i))]\} = 0, \tag{4}$$

where b is the estimator for β and $f(q)$ is the derivative of F with respect to β. This is exactly the same as the first-order condition for weighted non-linear least squares[14] with weights given by:

$$w_i = [F(bx_i)(1 - F(bx_i))]^{-\text{«}} \tag{5}$$

Using this notation, the estimate of the coefficient vector if the X matrix is known is given (when W is the diagonal matrix of w_i^2) by:

$$b = (X'WX)^{-1} X'WY \tag{6}$$

When the X matrix is unbiasedly imputed, the new estimator is:

$$b^* = (X^{*\prime}W^*X^*)^{-1} X^{*\prime}W^*Y \tag{7}$$

where the stars on the right hand side of (6) indicate that the missing values for the x_i have been unbiasedly imputed. By Slutsky's theorem (Judge *et al*., pp. 147-148), b^* will be a consistent estimator of β. Of course, the first order conditions are nonlinear and have to be solved using an appropriate algorithm such as Gauss-Newton or iteratively reweighted least squares. We conclude that replacing the missing observations with unbiasedly imputed observations does not affect the consistency of the ultimate estimator in a binary model.

c. Multiple Imputation of the Missing Data.

To compute the maximum likelihood estimator of the missing data points in the X matrix, some distribution of the variables in X must be assumed. Barring evidence to the contrary, it is reasonable to assume that the continuous data are multivariate normal. In fact, Little and Rubin (1987) have shown that assuming the data are multivariate normal is relatively robust, although some efficiency can be lost if the data come from another distribution. Wang, Sedransk, and Jinn (1992) show this method to be robust even when the data are not MAR. Little and Rubin (1987) show that consistency is retained when fully observed dummy variables are included in the data matrix (pp. 206-207). Finally, if the assumption of normality is very much at odds with the data, the algorithm we use will not converge (see Little (1992) and Little and Rubin (1987 chap. 10)).

The method of implementing the ML estimator is the Expectation-Maximization or EM[15] algorithm. In this application, the EM algorithm starts with an initial guess of the missing data points (zero is used for this guess) and computes (M step) the maximum likelihood estimate of the vector of means and the variance-covariance matrix of the x's using observed data and the estimated X matrix. Then, the conditional expectation of the missing data is computed (E step) using the available data, the estimated mean vector and the estimated variance covariance matrix. The steps are repeated until the difference between iterations in the estimated variances of the complete sets of data is below a predefined convergence criterion. (We specified this to be .001.) Since we assume ignorability of the missing data, the likelihoods evaluated are the likelihoods of the missing data, independent of the final model.

We assume the missing data are ignorable for two reasons. First, the null hypothesis is that none of the independent variables should have any effect at all on the independent variable. Under the null, any missing data must be ignorable. Second, the dependent variable and the data with missing values come from two completely different data sets. It is very difficult to construct a scenario under which the factors causing the data to be missing are related to the dependent variable at all, much less that the process causing the missing observations would effect the parameter estimates in our final model. The assumption of ignorability, along with the fact that all categorical variables are fully observed, allow us to estimate the missing data independent of the form of the regression.[16]

There are now at least two commercial software packages available for estimating missing data parameters using the EM algorithm. We used a Gauss (Aptech Systems 1988) add-on called MISS (Schoenberg 1995) and found both the program and support to be excellent. Little (1992) also references a BMDP (Dixon 1988) package. As stated above, the algorithm is computer intensive and convergence can be slow. Our data set took about 10 hours for each run on a Pentium-90 with 16 megabytes of memory. The EM algorithm is well-behaved, however. Little and Rubin (1987) show that each iteration increases the likelihood function (decreases the residual sum of squares) for exponential-family distributions with bounded parameters. They also show that EM converges to a stationary point and that convergence is linear with the rate of convergence proportional to the percentage of missing data.[17]

Little and Rubin (1987, chapters 7, 8, and 9) provide a full discussion of the EM algorithm,[18] and we follow their presentation here with minor changes in notation. The E step calculates (for n observations after the t'th iteration):

$$E[\textstyle\sum_i^n x_{ij} \mid (M^t, \Omega^t), X_{obs}] = \sum_i^n x_{ij}^t, \qquad \forall\ j, \tag{8a}$$

$$E[\textstyle\sum_i^n x_{ij}x_{ik} \mid (M^t, \Omega^t), X_{obs}] = \sum_i^n (x_{ij}^t x_{ik}^t + c_{jki}^t), \qquad \forall\ j,k, \tag{8b}$$

where

$$x_{ij}^t = x_{ij} \text{ if } x_{ij} \text{ is observed and} \tag{9a}$$
$$x_{ij}^t = E[x_{ij} \mid (M^t, \Omega^t), x_{obs,i}] \text{ if } x_{ij} \text{ is not observed,} \tag{9b}$$
$$c_{jki}^t = 0 \text{ if } x_{ij} \text{ or } x_{ik} \text{ are observed and} \tag{10a}$$
$$c_{jki}^t = Cov[x_{ij}x_{ik} \mid (M^t, \Omega^t), x_{obs,i}] \text{ if } x_{ij} \text{ and } x_{ik} \text{ are missing.} \tag{10b}$$

Here, M^t is the vector of means of the variables, Ω^t is the estimated variance-covariance matrix at the t'th iteration, the subscript *obs* signifies observed data, $x_{obs,i}$ represents the set of variables observed for observation i, and E is the expectations operator. The c_{jki}^t is added to equation (8b) because x_{ij}^t is a random variable if it is unobserved, and the expectation of the product of two random variables is equal to the product of their expectations plus their covariance. In equation (9b), $E[x_{ij} \mid (M^t, \Omega^t), x_{obs,i}]$ is the predicted value of x_{ij} from a regression of x_j on x_i^t (i not equal to j) using Ω^t and M^t to reconstruct the X matrix.

The M step calculates:

$$\mu_j^{t+1} = n^{-1} \textstyle\sum_i^n x_{ij}^t, \qquad \forall\ j, \tag{11}$$

$$\begin{aligned} \sigma_{jk}^{t+1} &= n^{-1} E[\textstyle\sum_i^n x_{ij}x_{ik} \mid M^{t+1}, X_{obs}] - (\mu_j^{t+1} \mu_k^{t+1}) \\ &= n^{-1} \textstyle\sum_i^n [(x_{ij}^t - \mu_j^{t+1})(x_{ik}^t - \mu_k^{t+1}) + c_{jki}^t], \qquad \forall\ j,k. \end{aligned} \tag{12}$$

This step just calculates the estimated mean vector and variance-covariance matrix from the estimated full data matrix that resulted from the E step. This information is then used for the regression in the next iteration.

After the missing data point parameters are consistently estimated, the missing values imputed and the ultimate regressions run, the question remains how the variances of the estimated coefficients are to be interpreted. If the imputed data points are conditional means (exactly on the regression line), the estimated coefficient variances provided by standard software packages will be understated and require adjustment. Additionally, Little (1992) shows that the coefficient estimates would not be consistent in that case. Little, Rubin, and others have derived the required adjustments for the variances under rather restrictive conditions, primarily when the data is missing in a block. When the missing observations are scattered throughout the independent variable matrix and are imputed with the maximum likelihood point estimate, calculating the required adjustments of the standard output variances is intractable. However, if the variance is corrected for at the time the X matrix is imputed, calculating the adjustments is easy. This method, called “multiple imputations” even when only one imputation is made, essentially involves imputing the variance of the data at the time the missing data is imputed.

With multiple imputations, described by Rubin in (1978), a pseudo-random number is imputed for the missing data point. The pseudo-random number is drawn from the distribution parameterized by the expected mean and variance of the missing observation, given the data and the maximum likelihood (EM) estimate of the mean and covariance matrices of the data set. This is repeated the desired number of times to create a new X

matrix with the same number of columns as the original, but with the number of rows equal to the product of the number of imputations and the number of rows in the original matrix.

This data set is then used exactly as a normal data set, and the coefficients need no adjustment. The variance of the coefficient estimate is $s\,(X'X)^{-1} = ((e'e)m^{-1})^{1/2}(X'X)^{-1}$. Because the X matrix is now the number of imputations times the original size, $m = nq$ where q is the number of imputations and n is the original number of observations. The standard errors of the coefficients can be interpreted normally after multiplying by the square root of the number of imputations. This method effectively weights each imputed data point by one over the number of imputations. Multiple imputations therefore, reduce the effect of outliers generated by the randomization of the estimated variance-covariance matrix without affecting the information in the known data points.

Immediately, the question of how many imputations should be done presents itself. Rubin and Schenker (1986) examine (analytically and using simulations) the effect of the number of imputations on the estimates of the coefficients and the estimates of the variances of the coefficients. They conclude that one imputation is unsatisfactory and that little is to be gained by more than two. We used one, two, and three imputations and compared the results. Our results agree with Rubin and Schenker. One imputation is superior to ignoring the observations containing missing data, if distance from the three imputations estimate is the criterion. Two imputations result in point and variance estimates very close to the three imputation case and appear adequate. This is useful because the size of a large data set such as ours multiplied by a large number of imputations might tax the resources of even a fast modern PC.

We wanted to compare this method's results with two methods that assume the data is MCAR, the complete case and a hot deck procedure. For the hot deck imputation, we used an extension of the simple method described in Section 2a. For each missing value we imputed draws from a normal distribution with the same mean and variance as the complete cases for that variable. The results of all of these procedures are reported in the next section.

IV. Results.

1. Definitions and Sources of Variables.

The variables used for the analysis were drawn from the Disclosure database (1993). This database consists of financial and nonfinancial data compiled from reports to the SEC. Since companies have different reporting dates, the data represent the most recent financial data received as of the Disclosure publication date. Companies are included if they provided direct goods and services and filed with the SEC. Therefore, the companies are publicly traded, listed over-the-counter or on a major exchange, have at least 500 shareholders, and have assets of at least $5 million. Companies that do not provide direct goods and services, primarily investment funds or management companies, are not included. In all cases except for the dummy variable indicating Green Lights membership, the variables were extracted from the Disclosure database without modification (except for changes of scale to make the coefficient estimates comparable in size). The Green Lights membership list was matched with the Disclosure data by company name, and the dummy variable created for Green Lights participants. The 1993

Disclosure data were used, with 1993 being a convenient midpoint between the starting date of the Green Lights program (January 1991) and the point in time at which membership in Green Lights was tallied (January 1995).[19] A subset of the Disclosure data base is provided by other vendors. This includes earnings estimates and forecasts provided by Zacks and ownership information provided by Spectrum. (See Disclosure (1994) for details.) Since these are third party data bases, they do not exactly match the coverage of the main Disclosure data base and are the source of many of the missing data points. Table 1 displays the variables, provides a brief description of each, and shows the percentage of the 9548 total observations that are missing. Table 2 presents the regression results.

{Insert Tables 1 and 2}

1. Discussion - Substantive Findings.

The first and strongest conclusion that emerges from the analysis is that the characteristics of firms do influence the probability of a company's joining the Green Lights program. Number of employees, earnings per share, the historical rate of growth of industry earnings, expected future earnings growth, the price/earnings ratio, a measure of insider control, industrial sector, and EPA region all have an influence on the discrete dependent variable.

The number of employees has a positive coefficient that is statistically significant at the .1% level. (The P-value for the two-tailed test of the hypothesis that the coefficient of employees is zero is less than .001.) The strong effect of this variable is probably associated at least in part with EPA's marketing effort---the Green Lights staff directed its efforts towards recruiting large companies first. The effect of size could also be associated with distributing the fixed costs of embarking on a lighting upgrade program over a wider base, although this argument is less compelling for large multi-divisional firms in which the operating units have considerable autonomy and act as independent profit centers.

Several of the statistically discernible effects are associated with performance characteristics of the firms. The coefficient of earnings per share (EPS) is significantly different from zero whatever the estimator, and the coefficient of the rate of growth of earnings per share in the firm's industry over the last five years (IEPSGL5) positive and strongly significant in all three cases with missing values imputed with EM. The coefficient of the price/earnings ratio (PE) is also positive and is statistically significant at the 5% level in the 3-imputation case. Both earnings per share and price/earnings ratio are indicative of good performance (and good expected future performance) of a firm. The price/earnings ratio may also be thought of as inversely related to the average cost of capital of the firm (under the simplifying assumption that the present earnings stream will continue indefinitely). This would suggest that firms are more likely to join Green Lights the lower is their average cost of capital. However, as argued in Section I above, the conventional theory of investment holds that the proper discount rate to use in evaluating these investments should be the same across all firms. Hence, the positive association between price/earnings ratio and the decision to join Green Lights is evidence against the conventional theory. It may be that firms with a lower cost of capital are more willing to make energy-saving lighting investments, but this finding supports the need to go beyond conventional investment theory to understand the behavior of these firms.

Interestingly, the industry's earnings per share growth rate performance appears to be more important than the company's own earnings per share growth rate as a determinant of the decision to join Green Lights. It could be, however, that the most recent earnings per share (EPS) is picking up the firm-specific earnings effect better than the historical 5-year EPS growth rate. The Zacks forecast of earnings growth over the next five years (GR5_NEXT) has a negative and significant coefficient. This variable may just be another size proxy (with smaller firms expected to grow more rapidly than large ones), or it could be that firms that have done well over the past five years are expected to do relatively less well over the next five. The Zacks forecast is an average of subjective forecasts, so it is difficult to know exactly what elements went into it; for purposes of our test it is not necessary to know those elements, because they clearly depend on characteristics of the firms and should be irrelevant under the conventional theory.

It should be noted that both the earnings per share variable (EPS) and the number of shares variable (SHAREBIL) include arbitrary scale factors. Actual performance should not be affected by a stock split, for example. Despite this arbitrary scaling, the coefficient of earnings per share is positive and statistically significant in all the regressions in which missing values are imputed.[20] SHAREBIL's coefficient is not statistically significant in any of the models in which the missing data are imputed by EM. Other measures of performance and capital structure (QUICK, NIOCE, and TDOE) do not show any statistically discernible influence on the decision to join Green Lights in the regressions with EM imputation. However, the measure of insider control (INSIDER1) has a negative and statistically significant coefficient. Under conventional theory, ownership structure should make no difference, yet an effect is clearly present.

Utilities (SIC 4900-4971) were more likely to join Green Lights, while finance, insurance, and real estate firms (SIC 6000-6799) and services companies (SIC 7000-8999) were less likely to join, relative to the omitted SIC dummy variable, which mainly consisted of manufacturing firms (SIC 2000-3999). (Transportation and communications (SIC 4000-4899) also had a negative coefficient significantly different from zero at the 5% level in the 3-imputation EM case.) In addition, companies headquartered in EPA Regions 9 (Arizona, California, Hawaii, Nevada, American Samoa, and Guam) and 10 (Alaska, Washington, Oregon, and Idaho) were more likely to join, as were companies located in EPA Regions 2 (New Jersey, New York, Virgin Islands, and Puerto Rico) and 3 (Delaware, District of Columbia, Maryland, Pennsylvania, Virginia, and West Virginia). The conventional theory would predict that none of these differences in sector or location should matter. But if joining Green Lights depends on diffusion of knowledge about the technologies, it is plausible that the utilities would be among the first to join. The observed regional effects could be the result either of differences in the level of effort by the regional EPA offices in implementing the program, of "epidemiological" effects arising from communications between firms located close to each other, or other locational factors influencing the diffusion process.[21]

2. Discussion - Comparison of Estimators.

Logit with three imputations by EM is obviously more efficient than logit with only the complete observations. The additional information allows us to conclude that several variables have coefficients that are significant at the 5% level (two tailed test) when all the information is used, but that these coefficients are indistinguishable from zero if only the complete observations are used. These variables include size as measured by employment (EMPLOYS), earnings per share (EPS), expected earnings growth over the next five years (GR5_NEXT), the price/earnings ratio (PE), and the dummy variables for transportation and communications (SICTRCO), utilities (SICUTIL), finance, insurance and real estate

(SICFIRE), services (SICSERV), and four of the EPA regions (EPAREG2, EPAREG3, EPAREG9, and EPAREG10). Also, the complete case regression indicates that the number of shares outstanding (SHAREBIL) has a coefficient that is statistically significant at the .001 level, but this variable's coefficient is not significant in the regressions in which the missing observations are imputed by EM. The coefficient of the price/earnings ratio (PE) is not significant and negative in the complete case, but it has a positive coefficient and a two-tailed P-value of .04 with 3-imputation EM. Similarly, the coefficient of the utility dummy is negative and non-significant in the complete case, but positive and strongly significant in all the EM-imputation estimates. Using only the complete observations would, in the case of these variables, lead to incorrect inference regarding both the signs and statistical significance of the coefficients.

The hot deck estimates are preferable to those based only on the complete observations, but these estimates also are misleading in some cases. While the signs of the hot deck estimates are usually the same as the signs of the EM estimates, hot deck shows a statistically significant (at the 5% level) estimate of the coefficient of the mining dummy and the quick ratio, while the EM coefficient estimates are not statistically significant for these variables. In addition, hot deck fails to reveal the influence of industry earnings growth (IEPSGL5). The hot deck estimates impute missing values using the correct unconditional means and variances of the variables, but do not take account the interdependencies among variables. Given the availability of practical methods for computing EM estimates, and the fact that inference hot deck requires more restrictive assumptions than inference under EM (MCAR, rather than ignorability), it would seem that hot deck's only virtue is computational ease.

3. Comparison to Other Studies.

Our findings are relevant to at least two streams in the literature. First, there are studies that examine directly the issue of participation in voluntary pollution prevention initiatives. These include papers such as those by Arora and Gangopadhyay (1995), Maxwell, Lyon, and Hackett (1995), Arora and Cason (1995), Winn (1994) and DeCanio (1994). Second, our findings contribute to the literature on why factors other than the cost of capital and the standard NPV calculations are important determinants of corporate investment. This literature is voluminous, and can be represented by the work of Fazzari, Hubbard, and Petersen (1988), Calomiris and Hubbard (1990), Hubbard and Kashyap (1992), Carpenter, Fazzari, and Petersen (1994), and Stole and Zwiebel (1996). We will not undertake any sort of review of this second body of literature; it is sufficient to note that our findings are consistent with the substantial current of economic thought holding that corporate investment behavior is considerably more complex than can be described by the bare-bones NPV model of investment.

Arora and Cason (1995) address the question of participation in the EPA's voluntary 33/50 program. This program is designed "to encourage firms voluntarily to reduce releases and transfers of 17 toxic chemicals" (p. 271). Arora and Cason carry out a logit regression analysis, with participation in the 33/50 program as the dependent variable and a variety of factors including sector, net income, total assets, employment, debt ratio, and indicators of the degree of competition in the firms' industries as explanatory variables. Their theoretical framework is a standard one in which firms decide to participate or not based on considerations of expected profit. Their analysis includes several variables similar to ours, and they also find that larger firms are more likely to participate in the program than smaller ones. Although the signs of their profitability and capital structure variables are as expected, the effects of these variables are not statistically significant and they conclude only that "the estimates...provide some weak evidence that capital markets constrain some firms from initiating environmental overcompliance" (p. 284). They also find significant industry effects. Unlike our case, however, this is to be expected under the conventional model because *the technologies of toxic waste abatement are very different across firms and sectors*. Indeed, Arora and Cason find a strong association between aggregate toxic releases and participation in the 33/50 program. This is consistent with technological factors being important determinants of participation.[22]

Unlike the pollution-control technologies employed by the Arora and Cason firms, technologies that vary according to the firms' products and processes, the Green Lights technologies are *generic and uniform* across firms and sectors. This is why our test can be focused more precisely on determining whether company differences that would not affect participation under a pure neoclassical economic model do in fact influence the decision to join. Our statistical methodology is also different, because of the nature of the data; the Arora and Cason sample was restricted to a subset of industries; their sample was considerably smaller (approximately 300 firms) than ours, and they did not take any measures to estimate missing data.

DeCanio (1994) examined whether there were any differences in performance between companies that joined Green Lights and others. The approach taken in that paper was to look at a matched pairs sample of Green Lights joiners and companies similar in size and product line. The data set was small (only about 40 matched pairs), which was compensated for in part by the increased power of matched pairs tests. DeCanio found that Green Lights companies did tend to have higher earnings growth rates than nonjoiners. This result is consistent with our results---companies showing positive indicators of performance are more likely to belong to Green Lights. The statistical methodology of our paper is superior, however, because (1) the data we are using is so much more extensive and (2) we avoid the inherently subjective process of choosing matched pairs.

Winn (1994) takes a very different approach. Her paper uses the multiple case study method to explore the factors influencing the adoption of pioneering corporate environmental policies. She found evidence for a number of organizational, cultural, and social influences that can be specific to particular firms but that are not captured at all by conventional economic profitability analysis. Winn saw evidence of the importance of “champions” of environmental investments and of their position(s) in the organizational hierarchy. She also found that outside forces mattered, some (such as publicity or consumer input) in ways that are consistent with the conventional analysis, but others (such as the role of environmental values held by management) that do not fit easily into the conventional framework. Winn’s organizational-theoretic approach is broadly consistent with our findings.

The papers by Arora and Gangopadhyay (1995) and Maxwell, Lyon, and Hackett (1995) examine theoretical models of or “voluntary overcompliance” with environmental regulation or “self-regulation” to reduce pollution. Both papers rest on profit-maximizing models of firm behavior, and thus can be seen as bridges between the two strands of literature referred to at the beginning of this section. These papers develop specific instances in which profit-maximizing firms would reduce pollution beyond what is required by law. In the model of Arora and Gangopadhyay, consumers value environmental quality but are different in their willingness to pay for it because of having different income levels. Firms distinguish themselves from each other by providing different levels of pollution control, resulting in a segmented market in which the firms selling to richer consumers reduce their emissions more than firms selling to poorer consumers. Maxwell, Lyon, and Hackett take an alternate approach. In their model, firms voluntarily self-regulate in order to preempt interest group pressures for regulation. As long as it is costly for consumers or environmentalists to organize, it is possible for firms to foreclose formal regulation by voluntary pollution abatement. While this abatement may be less than would be socially optimal, both firms and consumers benefit relative to the status quo when the costs of influencing policy are included in the calculation of consumers' welfare.

These papers have the virtue of showing how voluntary pollution prevention can be in firms' interests even under the conventional economic assumptions that pollution control is costly and that firms maximize profits. Our statistical results are a test only of the plain neoclassical model of investment, and as such they are not in conflict with the theoretical possibilities developed by Arora and Gangopadhyay and by Maxwell, Lyon, and Hackett.[23] However, it is important to bear in mind that there are other reasons the neoclassical investment hypothesis might fail. Firms are complex organizations, capable of deviations from profit maximization as well as of subtle adaptations to market opportunities such as those described in these recent papers. Further research is needed to determine how best to account theoretically for the empirical evidence pointing to failure of neoclassical investment theory.

V. CONCLUSIONS

Analysis of the factors influencing companies' decision to join the EPA's Green Lights program demonstrates that a wide range of company-specific characteristics are statistically associated with GL membership. The lighting retrofits covered by Green Lights are generic and the financial analysis of their potential profitability should not vary across firms, so the presence of these firm-specific influences is evidence that the conventional model of investment decision-making is inadequate in this case. Instead, it appears that Green Lights membership is positively associated with good performance by firms, and with sectoral and regional characteristics that suggest the importance of informational diffusion.

From a methodological perspective, it is clear that the use of the maximum likelihood EM/Multiple Imputation technique for analysis of data with missing values offers great practical advantages. It is possible to obtain substantial gains in estimation efficiency, as well as to avoid incorrect statistical inferences, by making use of all the data rather than only those parts of it that are "complete." Although the computational requirements are considerable, the maximum likelihood approach to missing data problems is well within the capability of personal computers. Our findings have corroborated the conclusions of other researchers on issues such as the robustness of maximum likelihood techniques and the required number of imputations if multiple imputations are used.

The substantive results of this study have implications both for corporate and government policy. Corporations have an opportunity to examine the degree to which their internal decision-making procedures might be reformed to improve profit performance. The usual preference for internal financing, if relaxed in the case of energy-efficiency investments of low risk, could enable the capturing of profit opportunities that might otherwise be missed. Government policy-makers also can benefit from the knowledge that private-sector firms are influenced to reduce pollution for reasons other than regulation. Government can provide incentives for action that are non-coercive and that promote improved corporate performance. By speeding the dissemination of information about energy-saving technologies, programs such as Green Lights can offer win-win benefits to taxpayers, consumers, and shareholders. Recognizing that internal characteristics of firms influence decision-making in ways not measured by conventional investment analysis can foster government policies that are more flexible than traditional tax or regulatory incentives. For economists and policy-makers, acknowledgment of the deeper structure of corporate decision making is an important hurdle that needs to be surmounted if the environmental problems of the 21st century are to be addressed by policies that go beyond those of the 20th.

REFERENCES

Antle, Rick, and Gary D. Eppen, 1985. "Capital Rationing and Organizational Slack in Capital Budgeting," *Management Science*, Vol. 31, No. 2 (February): 163-174.

Aptech Systems, Inc., Maple Valley WA, 1994. GAUSS manuals: Vol. 1 "System and Graphics Manual" and Vol. 2 "Command Reference."

Arora, Seema, and Timothy N. Cason, 1995. "An Experiment in Voluntary Environmental Regulation: Participation in EPA's 33/50 Program," *Journal of Environmental Economics and Management*, Vol. 28: 271-286.

Arora, Seema, and Subhashis Gangopadhyay, 1995. "Toward a theoretical model of voluntary overcompliance," *Journal of Economic Behavior and Organization*, Vol. 28: 289-309.

Ayers, Robert U., 1993. "On Economic Disequilibrium and Free Lunch," Centre for the Management of Environmental Resources Working Paper 93/45/EPS (INSEAD: Fontainebleau, France).

Bernanke, Ben S., 1983. "Irreversibility, Uncertainty, and Cyclical Investment," *The Quarterly Journal of Economics*, Vol. 97, No. 1: 85-106.

Calomiris, Charles W., and R. Glenn Hubbard 1990. "Firm Heterogeneity, Internal Finance, and 'Credit Rationing,'" *The Economic Journal*, Vol. 100 (March): 90-104.

Carpenter, Robert E., Steven M. Fazzari, and Bruce C. Petersen 1994. "Inventory Investment, Internal-Finance Fluctuations, and the Business Cycle," *Brookings Papers on Economic Activity*, Vol. 2: 75-138 (includes comments).

Cummins, Jason G., Kevin A. Hassett, and R. Glenn Hubbard 1994. "A Reconsideration of Investment Behavior Using Tax Reforms as Natural Experiments," *Brookings Papers on Economic Activity*, Vol. 2: 75-138 (includes comments).

Davidson, Russell and James G. MacKinnon, 1993. *Estimation and Inference in Econometrics* (New York: Oxford University Press).

DeCanio, Stephen J., 1993. "Barriers within firms to energy-efficient investments," *Energy Policy*, Vol. 21, No. 9 (September): 906-914.

______________, 1994. "Energy Efficiency and Managerial Performance: Improving Profitability While Reducing Greenhouse Gas Emissions," in *Global Climate Change and Public Policy*, ed. David L. Feldman (Chicago: Nelson-Hall Publishers): 86-101.

______________, 1995. "The Energy Paradox: Bureaucratic and Organizational Barriers to Profitable Energy-Saving Investments," Working Paper in Economics #19-95, Department of Economics, University of California, Santa Barbara.

Disclosure, Inc., 1993. "Corporate Information on Public Companies Filing with the SEC, Including Disclosure Spectrum" (Bethesda, MD: December).

______________,1994. *A Guide to Database Elements*. (Bethesda, MD: Disclosure, Inc.)

Dixon, W. J. (ed.), 1988. *BMDP Statistical Software Manual* (Vol. 2) (Berkeley, CA: University of California Press).

Dixit, Avinash, 1992. "Investment and Hysteresis," The Journal of Economic Perspectives, Vol. 6, No. 1: 107-32.

Dixit, Avinash K., and Robert S. Pindyck, 1994. *Investment Under Uncertainty* (Princeton: Princeton University Press).

Building Operating Management, 1992. "Guiding Lighting Retrofit," (March): I/2.

Fazzari, Steven M., R. Glenn Hubbard, and Bruce C. Petersen, 1988. "Financing Constraints and Corporate Investment," *Brookings Papers on Economic Activity*, eds. William C. Brainard and George L. Perry, Vol. 1: 141-195 (includes comments).

Greene, William H., 1993. *Econometric Analysis* (New York: Macmillan Publishing Company).

Griliches, Zvi, 1986. "Economic Data Issues, section 5: Missing Observations and Incomplete Data," *Handbook of Econometrics, Volume III*, eds. Z. Griliches and M. D. Intriligator: 1485-1495.

Hartley, H. O., 1958. "Maximum Likelihood Estimation From Incomplete Data," *Biometrics*, Vol. 14 (June): 174-194.

Hassett, Kevin A., and Gilbert E. Metcalf, 1993. "Energy conservation investment," *Energy Policy*, Vol. 21, No. 6 (June): 710-716.

Heitjan, Daniel F., and J. Richard Landis, 1994. "Assessing Secular Trends in Blood Pressure: A Multiple-Imputation Approach," *Journal of the American Statistical Association*, Vol. 89, No. 427 (September): 750-759.

Heitjan, Daniel F., and Roderick J. A. Little, 1991. "Multiple Imputation for the Fatal Accident Reporting System," *Applied Statistics*, Vol. 40, No. 1: 13-29.

Hubbard, R. Glen, and Anil K. Kashyap, 1992. "Internal Net Worth and the Investment Process: An Application to U.S. Agriculture," *Journal of Political Economy*, Vol. 100, No. 31: 506-534.

Huntington, Hillard, Lee Schipper, and Alan H. Sanstad, eds., 1994. "Markets for energy efficiency," special issue of *Energy Policy* (October).

Jaffe, Adam B., and Robert N. Stavins, 1993. "The Energy Paradox and the Diffusion of Conservation Technology," working paper (July 14).

Judge, George G., W. E. Griffiths, R. Carter Hill, Helmut Lutkepohl, and Tsoung-Chao Lee, 1985. *The Theory and Practice of Econometrics* (Second Edition). (New York: John Wiley and Sons).

Koomey, Jonathan G., 1990. "Energy Efficiency Choices in New Office Buildings: An Investigation of Market Failures and Corrective Policies," Ph.D. Dissertation, Energy and Resources, University of California at Berkeley.

Little, Roderick J. A., 1982. "Models for Nonresponse in Sample Surveys," *Journal of the American Statistical Association,* Vol. 77, No. 378 (June): 237-250.

________, 1992. "Regression With Missing X's: A Review," *Journal of the American Statistical Association,* Vol. 87, No. 420 (December): 1227-1237.

________, 1993. "Pattern-Mixture Models for Multivariate Incomplete Data," *Journal of the American Statistical Association,* Vol. 88, No. 421 (March): 125-134.

Little, Roderick J. A. and Donald B. Rubin, 1987. *Statistical Analysis With Missing Data* (New York: John Wiley & Sons, Inc.).

Lovins, Amory B., and L. Hunter Lovins, 1991. "Least-Cost Climatic Stabilization," *Annual Review of Energy and the Environment*, Vol. 16: 433-531.

Maddala, G. S., 1983. *Limited-Dependent and Qualitative Variables in Econometrics* (New York: Cambridge University Press).

Maxwell, John W., Thomas P. Lyon, and Steven C. Hackett, 1995. "Self-Regulation and Social Welfare: The Political Economy of Corporate Environmentalism," Indiana University Working Papers in Economics, Number 96-004. Bloomington, IN: Indiana University School of Business.

McKendrick, Lieut.-Col. A. G., 1926. "Applications of Mathematics to Medical Problems," *Proceedings of the Edinburgh Mathematical Society,* Vol. 44: 98-130.

Meng, Xaio-Li and Donald B. Rubin, 1991. "Using EM to Obtain Asymptotic Variance-Covariance Matrices: The SEM Algorithm," *Journal of the American Statistical Association*, Vol. 86, No. 416 (December): 899-909.

Pindyck, Robert S., 1991. "Irreversibility, Uncertainty, and Investment," *Journal of Economic Literature*, Vol. 29, No. 3: 1110-48.

Rubin, Donald B., 1976. "Inference and Missing Data," *Biometrika*, 63, 581-592.

Rubin, Donald B., 1987. *Multiple Imputation For Nonresponse in Surveys* (New York: John Wiley & Sons, Inc.).

Rubin, Donald B. and Nathaniel Schenker, 1986. "Multiple Imputation for Interval Estimation From Simple Random Samples With Ignorable Nonresponse" *Journal of the American Statistical Association*, Vol. 81, No. 394 (June): 366-374.

Ruud, Paul A., 1991 "Extensions of estimation methods using the EM algorithm," *Journal of Econometrics*, Vol. 49: 305-341.

Sanstad, Alan H., Carl Blumstein, and Steven E. Stoft, 1995. "How high are options values in energy-efficiency investments?" *Energy Policy*, Vol. 23, No. 9: 739-743.

Schoenberg, R. J., 1995. *MISS Reference Manual*. Seattle: RJS Software.

Stole, Lars A., and Jeffrey Zwiebel, 1996. "Organizational Design and Technology Choice under Intrafirm Bargaining," *American Economic Review*, Vol. 86, No. 1 (March): 195-222.

Tybout, James R., 1992. "Making Noisy Data Sing," *Journal of Econometrics,* Vol. 53: 25-44.

U.S. Environmental Protection Agency, 1995. "EPA Green Lights Program Snapshot: April 30, 1995" (Washington), May 16.

Wang, R., J. Sedransk, and J. H. Jinn, 1992. "Secondary Data Analysis When There Are Missing Observations," *Journal of the American Statistical Association*, Vol. 87, No. 420 (December): 952-961.

Winn, Monika, 1994. "Corporate Leadership and Policies for the Natural Environment." In: Post, J.E., Collins, D., and Starick, M. (Eds.), *Research in Corporate Social Performance and Policy: Sustaining the Natural Environment, Empirical Studies on the Interface between Nature and Organizations*. (Greenwich, CT: JAI Press): 127-161.

Figure 1 - Investment Opportunities of Two Firms

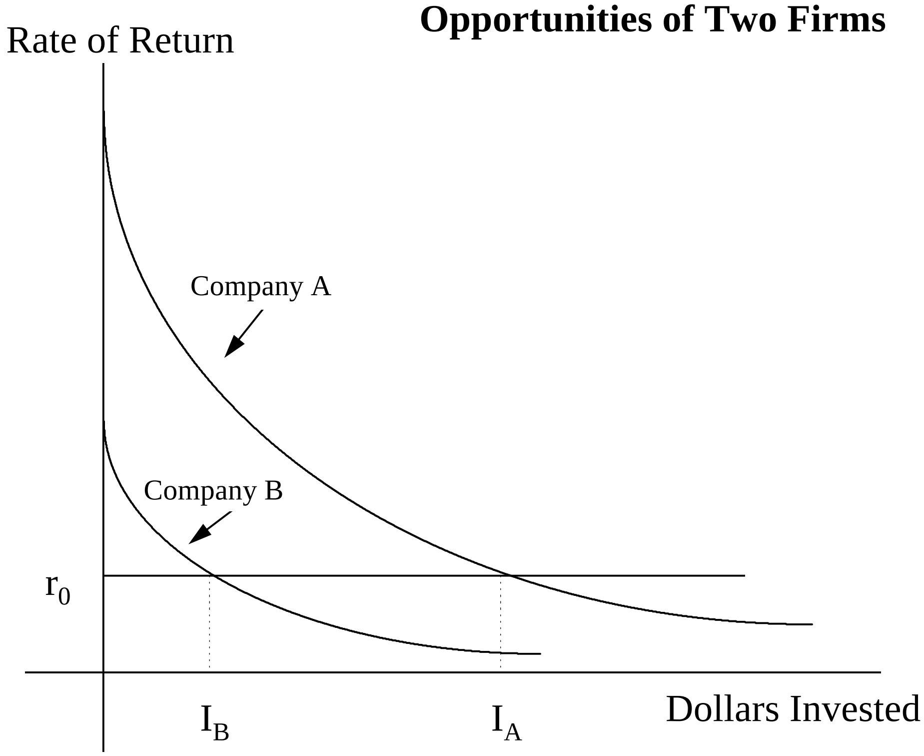

TABLE 1 - VARIABLE NAMES AND DESCRIPTIONS

VARIABLE	DESCRIPTION	% MISSING
GLJOIN	The dependent variable. It is a discrete variable, taking on the value 1 if the company had joined Green Lights as of January 1995, and 0 otherwise.	0%
EMPLOYS	Number of employees ÷ 1,000.	10%
EPS	Earnings for last 12 months per outstanding share.	39%
CEPSGL5	Company earnings-per-share growth rate over the last five years.	85%
IEPSGL5	Industry earnings-per-share growth rate over the last five years.	78%
SHAREBIL	Number of shares divided by one billion.	1%
GR5_NEXT	Forecasted earnings growth rate over the next five years.	68%
PE	(Closing price ÷ most recent 12-month's earnings)/100.	39%
QUICK	(Cash, marketable securities, and inventory ÷ current liabilities)/100.	6%
NIOCE	(Net income, after taxes ÷ company equity)/100.	4%
INSIDER1	(No. of shares owned by officers and directors) ÷ (No. of shares owned by officers and directors + No. of shares).	33%
TDOE	(Total debt to equity ratio)/100.	35%
SIC*	Dummy variable for SIC industry code, where the * indicates the industry. The default is manufacturing, while the included are: MINE, mining; TRCO, transportation and communications; UTIL, utilities; TRAD, trade; FIRE, finance, insurance and real estate; and SERV, services.	0%
EPAREG*	Dummy variable for EPA region. The * indicates regions 1 through 10 with 1 being the default.	0%

TABLE 2 - LOGIT RESULTS: Dependent variable is Green Lights membership

	(1) COMPLETE	(2) HOT DECK	(3) EM 1 IMPUTE	(4) EM 2 IMPUTE	(5) EM 3 IMPUTE
Log Likelihood	-316.136	-1020.953	-994.2106	-1000.668	-992.0607
C	-0.749510	-2.774346***	-2.653962***	-2.654319***	-2.590842***
	(.570838)	(.259158)	(.276258)	(.277795)	(.276125)
EMPLOYS	0.004812	0.010400***	0.008867***	0.009022***	0.008961***
	(.004074)	(.001584)	(.001651)	(.001638)	(.001651)
EPS	0.063736	0.106013***	0.155517***	.137197***	0.137896***
	(.062212)	(.023047)	(.029306)	(.028560)	(.029410)
CEPSGL5	-0.001641	-0.000773	-0.003439	-0.002787	-0.001292
	(.003541)	(.001629)	(.001784)	(.001812)	(.001871)
IEPSGL5	0.011217*	0.001925	0.008916***	0.007780***	0.010050***
	(.005431)	(.002101)	(.002381)	(.002345)	(.002392)
SHAREBIL	3.202213***	0.471916**	0.170742	0.079389	0.132050
	(.965941)	(.177948)	(.195454)	(.194944)	(.197449)
GR5_NEXT	-0.044018	-0.034610***	-0.048499***	-0.046979***	-0.052154***
	(.028890)	(.008834)	(.011273)	(.011213)	(.011295)
PE	-0.286776	0.087976	0.108704	0.102174	0.135171*
	(.204703)	(.0666580)	(.066474)	(.066420)	(.064512)
QUICK	-18.29483	-2.775232*	0.110413	-.815593	-0.003432
	(16.99307)	(1.145340)	(1.097817)	(1.181376)	(1.113035)
NIOCE	44.67062	-0.906152	-0.220929	-0.010826	-0.462437
	(35.55545)	(1.868853)	(1.347092)	(.848019)	(1.411966)
TDOE	-10.24118)	-0.191519	-0.338690	.061141	-0.032515
	(11.55373)	(.440014)	(.562968)	(.469147)	(.479016)
INSIDER1	-4.332581**	-3.783392***	-3.819743***	-3.913781***	-3.585565***
	(1.619149)	(.541673)	(.595534)	(.597775)	(.587534)

TABLE 2 (CONTINUED)

	(1)	(2)	(3)	(4)	(5)
SICMINE	-0.126687	-1.177258*	-0.783161	-0.787128	-0.709236
	(.883547)	(.473277)	(.485644)	(.485453)	(.486356)
SICTRCO	-0.809802	-0.592531	-0.619807	-0.652657	-0.670186*
	(.571523)	(.329427)	(.336211)	(.335285)	(.337177)
SICUTIL	-0.420786	1.035542***	0.664850***	0.608017**	0.606705**
	(.408490)	(.186216)	(.198453)	(.200241)	(.200216)
SICTRAD	-0.0464121	-0.440201	-0.326853	-0.339106	-0.368116
	(.404812)	(.249161)	(.250526)	(.250385)	(.251554)
SICFIRE	-0.326303	-0.955637***	-1.223683***	-1.230410***	-1.264775***
	(.321722)	(.201589)	(.208166)	(.207806)	(.209370)
SICSERV	-0.719926	-1.157985***	-1.067533***	-1.041451***	-1.067012***
	(.560073)	(.297658)	(.302296)	(.301870)	(.302584)
EPAREG2	-0.126493	0.579240*	0.689680**	0.728496**	0.735330**
	(.405307)	(.254483)	(.260753)	(.260606)	(.261340)
EPAREG3	0.045983	0.900013***	0.872550**	0.909294***	0.867916**
	(.417540)	(.265918)	(.273065)	(.273581)	(.274426)
EPAREG4	-0.579238	0.465622	0.560193*	0.581515*	0.550958
	(.461938)	(.278192)	(.284292)	(.285429)	(.286114)
EPAREG5	-0.682401	0.344862	0.363220	0.409911	0.350303
	(.386574)	(.253101)	(.257204)	(.257888)	(.258500)
EPAREG6	-0.218997	0.278502	0.474425	0.506758	0.495615
	(.486074)	(.295540)	(.300864)	(.301401)	(.301287)
EPAREG7	-1.264587	-0.144466	0.216937	0.202002	0.143391
	(.687960)	(.430929)	(.438983)	(.434647)	(.439487)
EPAREG8	-1.026915	-1.269402	-1.014494	-0.986438	-0.998203
	(1.091251)	(.740673)	(.742904)	(.743790)	(.743549)
EPAREG9	-0.598037	0.424754	0.676690*	0.662128**	0.637762**
	(.483284)	(.267965)	(.274404)	(.274735)	(.275178)
EPAREG10	0.728295	0.990882**	1.133631**	1.280972***	1.214940**
	(.564230)	(.372434)	(.376579)	(.376083)	(.378259)

Significance levels: * = 5%, ** = 1%, *** = .1%. Standard errors in parentheses below coefficients.

NOTES

1. Koomey (1990), Lovins and Lovins (1991), Ayers (1993), Jaffe and Stavins (1993), DeCanio (1993), and the special October 1994 issue of *Energy Policy* edited by Huntington et al. begin to scratch the surface of this literature. The area is ripe for a review article and/or meta-analysis.

2 . The EPA staff, of course, has an interest in “signing up” as many participants as possible for the programs.

3. For simple one-time investment expenditures such as these, the internal rate of return is well-behaved (does not have multiple roots), and the IRR is greater than the appropriate comparison discount rate if and only if the net present value of the project computed using that discount rate is positive.

4. This case is unlikely given the actual data on the profitability of lighting upgrade investments. Both engineering estimates and actual performance data suggest that the returns on straightforward lighting retrofits are far higher than any plausible estimate of the cost of capital for projects of comparable risk (DeCanio 1995). But we emphasize that this discrepancy in returns is not the basis of our test of conventional investment theory.

5. Originally, the profitability test was that the investments had to have internal rates of return (IRRs) greater than or equal to the prime rate plus six percent; recently, this requirement was changed to an IRR greater than or equal to a flat 20%. Since the projects are very low risk (proven technologies, little uncertainty in regulated electricity prices), the profitability hurdle was designed to guarantee that firms would only be committing to investments they should have wanted to make---under the neoclassical theory of investment---absent the Green Lights program.

6. While the data set includes categorical and continuous variables, missing data points are confined to independent continuous variables. We included categorical independent variables for SIC code and EPA Region of the company's headquarters, and these variables were complete for all observations.

7. Formally, data are missing completely at random (MCAR) if the distribution of the missing data indicator matrix $\mathbf{R}$, with $\mathbf{R}_{ij}=1$ if X_{ij} is observed and $\mathbf{R}_{ij}=0$ if X_{ij} is missing, does not depend on the observed or missing values of the data matrix; the data are missing at random (MAR) if the distribution of $\mathbf{R}$ depends on the data only through the observed values (Little 1992).

8. Rubin (1987) provides perhaps a more intuitive definition of ignorable: "ignorable...means that a nonrespondent is just like a respondent with the same value of *X*" (p. 22).

9. Consider the simple example of missing observations on the size of a firm. If they are missing because the management position responsible for reporting the data was not filled when the data were collected and the position openings were independent of any characteristics of the firm, the data are MCAR. If West-Coast firms were more likely to have a this management slot unfilled, the data are MAR. If all West-Coast firms, conditional on the other variables, have the same effect on the dependent variable whether the open management position is filled or not, the missing data are ignorable and MAR.

10. Only a small percentage of firms in the data set (less than 3%) joined Green Lights. This makes it important to estimate the parameters efficiently; a higher proportion of Green Lights members would improve the precision of estimators based on ordinary techniques.

11. Stratifying can be done by any independent variable. Examples in our case might include size of firm or industry.

12. Multiple and fine stratifications may approach the method we use, but our method is inherently superior because it uses all information in the data set to estimate the missing data points.

13. The derivation in this paragraph relies on material from Davidson and MacKinnon (1993, pp. 512-528), augmented with detail from Greene (1993) and Maddala (1983).

14. A general unweighted non-linear least squares estimator is obtained from

$$y = H(\$x) + e$$

by solving the first order condition

$$0 = \text{G}\,[H(bx) - y]\; h(bx)x.$$

The form of the weighted non-linear least squares estimator follows immediately.

15. Although not named until 1977 by Dempster, Laird, and Rubin, the EM algorithm has been used in limited applications for many years. Since it is a maximum likelihood method of estimating parameters, it has applications besides missing data estimation. Lieutenant Colonel A. G. McKendrick of the Laboratory of the Royal College of Physicians, Edinburgh, first introduced a version of the EM algorithm in 1926 to identify transmission mechanisms of diseases. He estimated the distribution of an epidemic under different assumptions and then matched the estimated distributions with observed patterns. He showed that this method allowed the investigator to distinguish agents of the transmission of the disease, such as a contaminated wells, fleas, or mosquitoes (McKendrick 1926). In

the first published use of EM to estimate missing data, H. O. Hartley in *Biometrics* (1958) proposed the algorithm and provided examples using Poisson and Binomial distributions. More recently, it has found a variety of applications including time series smoothing and factor analysis (Ruud 1991). Since the procedure can be slow to converge and is extremely computer intensive, little work was done to develop EM prior to the mid 1970's. Since then, Rubin and Little and their students have made substantial progress on the theoretical properties of the various estimators of missing data, and software has been written than enables the EM algorithm to be used as a practical tool for applied researchers.

16 . Little and Rubin (1987, chapter 10) show by an application of the factored likelihood theorem that estimating ignorable missing continuous data in the presence of catagorical data, when the continuous variables are the only variables with missing observations, is consistently done independent of the final regression model using the algorithm described in the text. This is because under normality the vector of means and the $X'X$ matrix are sufficient statistics for the joint distribution of the continuous variables. If the missing data were not ignorable, it would be necessary to use a different algorithm that would simultaneously estimate the coefficients of the logit model and the missing values.

17. Or, as an anonymous referee said "... the good behavior of the quasi-likelihood approach is at work here, and the convergence of the E-M algorithm is assured even if the

distribution of the *X*'s is not truly normal, provided that the true distribution belongs to the exponential family of distributions."

18. Here and in the remainder of the discussion of how the EM algorithm works, the *"X"* matrix is understood to include both dependent *and* independent variables. Little (1992) has pointed out that excluding the dependent variable when imputing the missing data points biases can bias the ultimate parameter estimates. Thus, even though none of the observations on the dependent variable is missing in our application, we include the dependent variable in the imputation process.

19. The profits earned on Green Lights investments were small relative to firms' total earnings, and so are not likely to be providing any discernable feedback on the explanatory performance variables. It typically takes a company some time after joining Green Lights to see the effects, because the first step is surveying facilities and deciding what investments to undertake. Also, many Green Lights partners initially undertake small pilot projects to see that they are satisfied with the results, to test employee responses to the lighting upgrades, etc. For these reasons, as well as the fact that some of the Disclosure information was collected prior to the Disclosure publication date, we do not believe simultaneity due to overlap between the time the performance variables were measured and the time companies joined Green Lights is a problem. Using Disclosure data from an earlier publication might weaken the test because the data would be obsolete for the period when companies were joining Green Lights.

20. Of course, the growth rate of earnings per share is not affected by the scaling.

21. Note that EPA Region 10 (except for Alaska) has the lowest electricity costs of the nation while Regions 2 and 3 has relatively high electricity costs. We did not include the price of electricity as an explanatory variable because the actual operations of multi-division firms are spread across several states, so that the price of electricity at the headquarters does not necessarily reflect the electricity price at the plants or offices where the investment would actually be made. Also, as shown in Section I, the electricity price, while surely influencing the number and profitability of projects a company would undertake, should not be a factor in determining whether or not the firm joins Green Lights. Trial regressions not reported here suggest that state-level electricity prices tended to have a negative influence on the decision to join GL, taking account of the other variables in our regressions. However, we suspect that this type of "price effect" is spurious, and is in actuality a representation of some other regional effect not being measured by the other variables. Other firm characteristics still show up as statistically significant influences on the decision to join Green Lights in these regressions, so inclusion or exclusion of price variables does not affect our main conclusions.

22. Arora and Cason also find some evidence that participation is more likely in less concentrated industries.

23. It is interesting to note that Maxwell, Lyon, and Hackett suggest, as one of the possible extensions of their model, that “large firms might undertake voluntary abatement while smaller firms choose to take a free ride on their rivals’ efforts” (ibid., p. 24). This could account for our finding that Green Lights membership is positively associated with firm size.