Modelling the consumer’s decision to replace durable goods: a hazard function approach

JENNIE E. RAYMOND, T. RANDOLPH BEARD and DANIEL M. GROPPER

Department of Economics, College of Business, Auburn University, Auburn, AL 36849, USA

This article analyses the consumer’s durable good replacement decision using hazard models. In contrast to the typical limited dependent variable model often used in durable good demand studies, hazard models allow for much richer relationships between the ages of durable goods and the probabilities of their replacement. To illustrate the technique, a recursive system consisting of a regression equation and a hazard model is used to examine home heating system replacement decisions by residential customers of a major southeastern US electric utility. The results indicate that overall system replacement rates decline over time, and that the probability of replacement for specific households depends negatively on the age of the head of household and the availability of natural gas, and positively on system age and higher than expected household energy use.

I. INTRODUCTION

Statistical analysis of the demand for durable goods is complicated by several factors. First, by virtue of their extended effective lifetimes, the acquisition and replacement of durable goods involves an element of timing which does not arise in typical demand studies. As durable goods age, the consumer is faced with the issue of deciding when (and if) to replace them, and such actions are often undertaken prior to irreparable failure. Hence, the timing of the replacement decision is critical. Second, because of their longevity, the typical consumer replaces products such as central space heating systems, refrigerators and automobiles infrequently, leading to some data-driven difficulties in analysing durable goods purchases with many conventional statistical techniques. Finally, since the demand for many durables is, like the demand for energy, a derived demand, numerous factors which play an important role in durable good purchase behaviour, such as consumer tastes for ‘comfort’, are impossible to observe directly.

This paper demonstrates an approach for the empirical study of durable replacement activity that addresses all three of the complicating factors mentioned above. This approach, which is based on the estimation of hazard (duration) models for durable goods replacement, offers a different and theoretically accommodating way to analyse durable good replacement decisions without reliance on logit, probit or other inherently static specifications. The procedure is illustrated by analysing the replacement of home heating systems among customers of a large US electric utility. The estimated results are useful in several ways. First, information is obtained on the importance of appliance age, energy use characteristics and other factors in the durable good replacement decision. Second, the results provide a means to estimate replacement rates over a wide range of time periods. Further, while the application focuses on home heating systems, the technique could be applied to many energy consuming durable goods.

II. PREVIOUS RESEARCH

Economists have long noted the value of distinguishing between the demand for a stock of durable goods, the
purchase of a durable good, and demand for the services provided by durables. Depending on the goal of the estimation, one or another of these features comes to the fore. As the primary interest of the present study is the decision to replace durable goods, the discussion will focus primarily on the purchase decision itself, and how this has been modelled previously.\footnote{See e.g. Philips (1987).}

Because consumers typically buy only one unit at a time of most durable goods, analyses of consumer purchase decisions on the individual level have usually utilized some type of limited dependent variable (LDV) statistical technique, typically probit or logit models.\footnote{Acquisition patterns for durable goods have been examined by Kasulks et al. (1979) and by Dickson et al. (1983). Consumer durable replacement decisions and the timing of replacement have been studied by Bayus (1991). Bayus and Carlstrom (1990) have analysed ways in which consumer durables can be grouped in order to better model customer preferences and purchase decisions.} The consumer's purchase decision is assumed to depend on various factors that economic theory suggests might influence the consumer's choice. Using the three part distinction mentioned above, these factors might include relevant prices, consumers' incomes, consumer taste variables and, in the case of derived demand, factors influencing the value of the services which the durable good will give the consumer.

The consumer's choice to replace an existing unit may also be influenced by the potential energy savings obtainable with a replacement. Since repairing and maintaining an older unit is often a plausible alternative to acquiring a new unit, the age of the unit to be replaced is obviously relevant. Although appliance age can be included in a typical LDV-type model, such a formulation is inherently restrictive. Since LDV models are useful only to predict replacements in the 'next' time period, they are of limited usefulness as models of complex dynamic processes.

### III. HAZARD MODEL ESTIMATION OF DURABLE GOODS REPLACEMENT

Hazard models are dynamic, non-linear statistical models which can be used to estimate the effects of observable characteristics on the length of time until a discrete event occurs. While many early hazard estimations focused on medical, biological or engineering problems, the first use of a hazard model in economic analysis was by Lancaster (1979), who studied unemployment duration. Amemiya (1985), Kiefer (1988) and Lancaster (1990) provide comprehensive surveys of this approach.

Formally, a hazard model assumes that the length of time, $t$, until an event occurs is a random variable with density $f(t)$ and cumulative distribution $F(t)$. Two auxiliary functions are derived from the density: the survivor function, $1 - F(t)$, which shows the probability that the event has not occurred by some particular time, and the hazard rate. The hazard rate, $h(t)$, is the likelihood that an event occurs at a particular time given that it has not occurred previously: $h(t) = -f(t)/(1 - F(t))$.

In practice, the hazard rate is specified initially and the density and survivor function are derived from it. The hazard rate may be hypothesized to depend on observable exogenous variables and on time itself. The dependence of the hazard rate on time implies that, as time passes and the event does not occur, the probability that the event occurs changes. The flexibility with which hazard models handle the impact of time on probability constitutes a major advantage over typical LDV formulations.

Here we hypothesize a hazard function of the following form:

$$h(t, x) = \exp(\gamma t) \cdot \exp(\beta' x) \quad (1)$$

where $x$ is a vector of exogenous variables, $t$ is time, and $\gamma$ and $\beta$ are parameters to be estimated. The sign of $\gamma$ will indicate whether the probability of replacement increases or decreases over time.\footnote{With multiple purchases, static approaches include the use of Poisson regression techniques (see Paull, 1978). The use of discrete choice models was pioneered by Farrell (1954). Other pioneering papers include Cramer (1962), Wu (1965) and Cragg (1971). This disaggregated approach is to be distinguished from aggregate analyses of durable good purchases, such as the approach of Chow (1960). Deaton and Muellbauer (1988) offer an extensive overview of theory and practice in this area.}

The expression for the hazard given in Equation 1 allows the derivation of the conditional distribution of time to failure, $f(t, x)$, by the relation: $h(t, x) = f(t, x)/(1 - F(t, x))$. We note that

$$f(t, x) = h(t, x) \cdot (1 - F(t, x)) \quad (2)$$

and that

$$1 - F(t, x) = \exp \left(- \int h(u, x) \, du \right) \quad (3)$$

The likelihood function is composed of two types of terms, the first type corresponding to those observations for

$h(t, x) = \exp(\gamma t + \beta' x) \cdot \exp(\beta' x)$

This specification was also estimated, but the results were not significantly different from those reported here. These results are available from the authors upon request.
which the hazard was observed to occur (‘completed spells’), and
the second representing consumers observed not to have
experienced the hazard. By simple substitution, the
likelihood function written in terms of the hazard rate is
\[
L = \prod_{i=1}^{N_1} h(t_i, x_i) \cdot \exp \left\{ - \int h(u_i, x_i) \, du_i \right\} \cdot \prod_{i=1}^{N_2} \exp \left\{ - \int h(u_i, x_i) \, du_i \right\}
\]  
(4)
where \(N_1\) and \(N_2\) are the number of complete and incom-
plete spells, respectively.

Having broadly outlined the estimation strategy, some
modifications of this approach are considered which are
made necessary both by the economic nature of consumer
behaviour and by some special features of the data set.

The application involves space heating system replace-
ment by a randomly selected sample of households buying
electricity from Alabama Power Company, a large regional
US utility. These households provided extensive information
on energy usage and household characteristics as part of
Alabama Power Company’s 1990 Residential Customer
Survey Program. In addition, the households’ energy usage
and billing records were made available to us in connection
with this survey.

All customers in our sample had some form of home
heating system during the survey period. Hence, all con-
sumer purchases of heating systems during the sample
period represented replacements. In addition to replacements
due to irreparable failures of existing systems, it seems
plausible that such replacements arose from two broad
motives: reliability and performance advantages rep-
resented by new units and potentially great energy savings
from modernization. This second motive makes obvious
a conceptual issue suggested by the derived nature of heat-
ing system demand: the value to a household of replacing an
older unit depends on individual tastes for energy and the
services (e.g. warmth or comfort) energy provides. House-
holds which have strong preferences for appliance services
would value improvements in efficiency more highly, and
would therefore be more likely to replace an older unit with
a new one.

The logic outlined above suggests how households’ ‘un-
observable’ tastes for energy services could be incorporated
into the estimation. Since a households’ energy usage de-
pends on both their tastes for energy services and their stock
of energy consuming durables, and since all sample house-
holds face identical electricity tariffs, similar geographic
circumstances, and highly similar appliance prices, observed
household energy usage rates, when adjusted for household
appliance stocks, offer an attractive way to quantify differ-
ences in tastes for appliance services and, by implication, the
values households place on improved efficiency.

To include a demand intensity measure into our estima-
tion, the following approach is taken. Each consumer’s
average monthly usage of electricity (over an end-of-sample,
one-year period) is modelled as a linear function of the
household’s stock of energy using durables and exogenous
factors, \(Z\)
\[
U = \theta Z + \epsilon
\]  
(5)
where \(U\) is average usage in kwh/month, \(\theta\) is a vector of
parameters to be estimated, and \(\epsilon\) is a normal disturbance
which is interpreted as representing idiosyncratic tastes. Let
\(\hat{U}\) be the predicted value for \(U\) and let \(\hat{v} = U - \hat{U}\) be the
estimated residual. Then the intensity variable, \(v\), is defined
as
\[
v = (U - \hat{U}) / \hat{U} = \epsilon / \hat{U}
\]
Thus, \(v\) provides a measure of the percentage deviation in
customer usage unexplained by house size, the stock of
durables owned and other observable factors. Computed for
each consumer individually, these demand intensity
measures can be used as explanatory variables, thereby
controlling for unobservable individual tastes.

Additionally, the nature of our data set requires a modifi-
cation to account for the large differential in the quantity of
information on appliance ages between customers who have
replaced their heating systems and those who have not. In
particular, for those who did not replace their systems in the
3 years prior to the survey date, information was obtained
on the age of their current system, but none on the age of
their previous system which, for some individuals, had been
replaced. For those who did replace their heating system in
those 3 years, the age of the replaced heating system was
noted. Since the sample of ages of systems at replacement
represents a random sample of the ages of all heaters at
replacement, the likelihood function was modified by
weighting observations for those who replaced their systems
by the inverse probability of replacement, and then nor-
malizing all weights to sum to one.\(^5\)

IV. MODEL ESTIMATION AND RESULTS

Our sample was taken from the Alabama Power Company
(1990), Residential Customer Survey and Alabama Power
Company billing records. Only those individuals who own
their homes, and for whom the home was the primary
residence, are included in the hazard model estimation (534
out of the original 702 observations).

To estimate \(v\), the demand intensity variable, average
kwh/month are regressed on a set of variables designed to
capture a variety of household characteristics. These vari-
ables are defined in Table 1, which presents the estimation
results for the demand intensity regression. These results are

\(^5\) The probability of replacement within the past 3 years is 11%, so these observations have an unnormalized weight of about 9.
Table 1. Household electrical energy use (average kWh/month) regression results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>256.16</td>
<td>0.77</td>
</tr>
<tr>
<td>Water heater age (years)</td>
<td>-6.47</td>
<td>-1.35</td>
</tr>
<tr>
<td>Space heater age (years)</td>
<td>7.83*</td>
<td>2.11</td>
</tr>
<tr>
<td>Number of electric heat pumps</td>
<td>20.41</td>
<td>0.23</td>
</tr>
<tr>
<td>Number of electric central systems</td>
<td>142.63*</td>
<td>1.82</td>
</tr>
<tr>
<td>Number of electric window AC units</td>
<td>22.48</td>
<td>0.42</td>
</tr>
<tr>
<td>Number of gas central systems</td>
<td>-77.19</td>
<td>-0.25</td>
</tr>
<tr>
<td>Average age electric heat pumps (years)</td>
<td>41.85*</td>
<td>7.97</td>
</tr>
<tr>
<td>Average age electric central system (years)</td>
<td>3.95</td>
<td>0.59</td>
</tr>
<tr>
<td>Average age electric window units (years)</td>
<td>0.59</td>
<td>0.12</td>
</tr>
<tr>
<td>Average age gas central system (years)</td>
<td>-8.48</td>
<td>-0.22</td>
</tr>
<tr>
<td>House square footage (1000s)</td>
<td>0.07</td>
<td>1.64</td>
</tr>
<tr>
<td>Single family home dummy</td>
<td>-278.17*</td>
<td>-2.30</td>
</tr>
<tr>
<td>Duplex dummy</td>
<td>282.56</td>
<td>1.15</td>
</tr>
<tr>
<td>Tri- or quadruplex dummy</td>
<td>349.90</td>
<td>1.43</td>
</tr>
<tr>
<td>5–9 family home dummy</td>
<td>-237.07</td>
<td>-0.81</td>
</tr>
<tr>
<td>10+ family home dummy</td>
<td>-192.23</td>
<td>-0.85</td>
</tr>
<tr>
<td>Urban/rural dummy (=1 if urban)</td>
<td>-11.60</td>
<td>-0.19</td>
</tr>
<tr>
<td>Number living in home</td>
<td>100.86*</td>
<td>4.79</td>
</tr>
<tr>
<td>Resident owns home dummy</td>
<td>62.35</td>
<td>0.62</td>
</tr>
<tr>
<td>Natural gas water heater dummy</td>
<td>-278.82</td>
<td>-0.90</td>
</tr>
<tr>
<td>Electric water heater dummy</td>
<td>92.87</td>
<td>0.30</td>
</tr>
<tr>
<td>Bottled gas water heater dummy</td>
<td>-280.49</td>
<td>-0.86</td>
</tr>
<tr>
<td>Number of type 1 appliances</td>
<td>163.47*</td>
<td>5.44</td>
</tr>
<tr>
<td>Number of type 2 appliances</td>
<td>33.27*</td>
<td>2.85</td>
</tr>
</tbody>
</table>

N = 702 observations  
$R^2 = 0.3168$  
F = 13.098

*Indicates significance at the 5% level. The omitted housing type is mobile homes. Type 1 appliances are dishwashers, clothes washers and dryers. Type 2 appliances are televisions, stereos, refrigerators and VCRs. The dependent variable is average kWh usage per month over a 1-year period immediately prior to the survey.

relatively intuitive and straightforward. It is noted that square footage and the numbers and ages of appliances typically have the expected effects on usage: big houses with older appliances use more energy. Additional significant effects are represented by heating system ages, multiple heating systems, the number of residents in the household, the housing type and the numbers of various appliances owned.

In the specification of the hazard function, the exogenous variables in addition to a constant term are the demand intensity variable $v$, the age of the head of the household, household income, a dummy variable indicating that the customer lives in an urban area, square footage of the home, a dummy variable indicating that natural gas service is available and a dummy variable indicating that the customer has a poor credit rating.\(^6\)

Table 2 presents the results of the hazard model estimation using the exponential specification with linear time-dependence as outlined in Section III. While these estimates are used later in some illustrative calculations, it is worthwhile to examine these results in some detail.

It is first noted that, unsurprisingly, the passage of time is highly significant in explaining replacement behaviour: the older the system, the greater the probability that it will be replaced. Further, the larger the value of the demand intensity variable, the more likely is home heating system replacement, a result consistent with the theoretical prediction. Two other factors are statistically significant indicators of replacement: the age of the head of the household, and the availability of natural gas service. The older the head of the household, the less likely he or she is to replace the heating system. Natural gas availability is also associated with a significantly lower probability of obtaining a new heating system, a result that may stem from possible differentials.

\(^6\) Households not reporting system replacement in the previous 3 years were treated as incomplete spells. Most of the variables included in the hazard function are largely self-explanatory. While the energy intensity variable has been described above, the poor credit rating dummy variable was taken directly from the utility company's rating system. All customers face the same electricity price schedule and, living in the same region, presumably face similar appliance prices except for urban/rural differentials.
Table 2. Hazard model coefficients

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.29*</td>
<td>-1.77</td>
</tr>
<tr>
<td>$v$: unexplained energy usage (per 10 percentage points)</td>
<td>0.14*</td>
<td>1.89</td>
</tr>
<tr>
<td>Age of head of household (per 10 years)</td>
<td>-0.13**</td>
<td>-4.41</td>
</tr>
<tr>
<td>Income (per $10,000)</td>
<td>-0.02</td>
<td>-0.76</td>
</tr>
<tr>
<td>Urban/rural dummy (=1 if urban)</td>
<td>-0.16</td>
<td>-1.46</td>
</tr>
<tr>
<td>House square footage (per 1000)</td>
<td>-0.01</td>
<td>-1.51</td>
</tr>
<tr>
<td>Natural gas availability dummy</td>
<td>-0.34**</td>
<td>-3.31</td>
</tr>
<tr>
<td>Poor credit rating dummy</td>
<td>-0.22</td>
<td>-1.47</td>
</tr>
<tr>
<td>Time (years)</td>
<td>0.28**</td>
<td>7.26</td>
</tr>
</tbody>
</table>

$N = 534$ observations

Note: Positive (negative) coefficients imply decreases (increases) in replacement time. t-values are asymptotic.
* Indicates significance at the 10% level.
** Indicates significance at the 1% level.

Table 3. Marginal impacts on the probability of replacing the primary home heating system

<table>
<thead>
<tr>
<th>Variable</th>
<th>1–3</th>
<th>4–6</th>
<th>7–9</th>
<th>1–20</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v$ (change is 10 percentage points)</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.004</td>
</tr>
<tr>
<td>Age of head of household (change is 10 years)</td>
<td>-0.009</td>
<td>-0.008</td>
<td>-0.007</td>
<td>-0.041</td>
</tr>
<tr>
<td>Income (change is $10,000 in 1990 dollars)</td>
<td>-0.002</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.007</td>
</tr>
<tr>
<td>Urban</td>
<td>-0.011</td>
<td>-0.010</td>
<td>-0.008</td>
<td>-0.050</td>
</tr>
<tr>
<td>Square feet (change is 1000 sq. ft)</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.004</td>
</tr>
<tr>
<td>Natural gas available</td>
<td>-0.024</td>
<td>-0.020</td>
<td>-0.017</td>
<td>-0.010</td>
</tr>
<tr>
<td>Poor credit rating</td>
<td>-0.016</td>
<td>-0.013</td>
<td>-0.011</td>
<td>-0.069</td>
</tr>
<tr>
<td>Overall probability of replacement</td>
<td>0.073</td>
<td>0.068</td>
<td>0.063</td>
<td>0.402</td>
</tr>
</tbody>
</table>

$T$-values are in parentheses. Marginal impacts are evaluated at sample mean values.

in the effective lifetimes of gas versus electric powered systems.\(^7\)

The advantage of hazard estimation over LDV procedures can be illustrated by making the following simple calculations. First, consider the marginal impacts of the regression variables on the probabilities of system replacement, calculated as the change in probability of replacement (over a specified future interval) per unit change in the explanatory factor. Table 3 presents these results for various time intervals. For example, a 10 year increase in the age of the head of the household reduces the probability of replacement within 20 years by a little over 4%, a highly significant effect. Other significant effects include demand intensity, with higher intensity increasing replacement probabilities, and natural gas availability, which reduces the likelihood of replacement by about 10% within 20 years.

Table 4 offers a similar analysis focused on the expected lifetimes of home heating systems, and how these estimated times until replacement are affected by changes in exogenous variables. Overall, the mean time to system replacement is 25 years.

However, high energy usage intensity reduces the expected lifetime, with a 10% increase in unexplained usage shortening time to replacement by almost 2 years. Similar

\(^7\) Replacement systems may be either gas or electric powered. However, those without gas service would always replace an electric powered system.
Table 4. Marginal impacts on the expected time to replacement of the primary home heating system (years)

<table>
<thead>
<tr>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>v (change is 10 percentage points)</td>
</tr>
<tr>
<td>Age of head of household (change is 10 years)</td>
</tr>
<tr>
<td>Income (change is $10000 in 1990 dollars)</td>
</tr>
<tr>
<td>Urban</td>
</tr>
<tr>
<td>Square feet (change is 1000)</td>
</tr>
<tr>
<td>Natural gas available</td>
</tr>
<tr>
<td>Poor credit rating</td>
</tr>
<tr>
<td>Overall time to replacement (years)</td>
</tr>
</tbody>
</table>

Marginal impacts are evaluated at sample mean values.

intuitively appealing effects are noted for the other explanatory variables.

V. SUMMARY AND CONCLUSIONS

This paper introduces the statistical technique of hazard model estimation to the study of durable goods replacement behaviour. The results obtained in the hazard model estimation are intuitively plausible and of statistical significance. While older systems are significantly more likely to be replaced, older heads of household are significantly less likely to obtain new systems. The availability of natural gas service also significantly reduces the probability of replacement. Further, those households with electricity consumption in excess of predicted levels are significantly more likely to replace their systems even when the age and type of their current system(s) is taken into account. This result conforms to the expectations of economic theory regarding durable goods replacement, and is important in explaining replacement behaviour.

While data limitations required us to limit our analysis to home heating systems, the methodology and techniques developed here are potentially applicable to a wide variety of durable good acquisition and replacement issues. While virtually all individuals in our sample had some form of home heating system, so that any purchase of a new system was a ‘replacement’, the initial (non-replacement) acquisition of other, ‘non-saturated’ appliances could also be studied with the hazard methodology. In this way, saturation issues, which are often of interest to energy analysts, can be directly evaluated with the hazard model approach. Hazard models also allow the researcher to overcome the inherently static approach of limited dependent variable modelling techniques. Our results show that the overall probability of replacement and the marginal impact of changes in the independent variables on the probability of replacement change over time. These findings would have been unobservable with traditional limited dependent variable techniques. Thus, it is concluded that the hazard model approach outlined and applied in this paper provides a useful and complementary alternative to the techniques traditionally used to model consumer behaviour in replacing durable goods.

ACKNOWLEDGEMENTS

Funding support from the Auburn Utilities Research Center is gratefully acknowledged. We would like to thank Jim Anderson, Andy Barnett, Hugh Granade, Gibson Lanier, John McDavid and an anonymous referee for helpful comments, and Bob Bolen, Dave Townley, Marty Wade and Bob Wenger for help with the data. The usual caveat applies.

REFERENCES