Price Dispersion and the Ability to Search: 

The UK Internet Motor Insurance Market

Stephen McDonald and Colin Wren

Newcastle University Business School, NE1 7RU, Newcastle upon Tyne, UK.

October 2009

No 2009/07

Newcastle Discussion Papers in Economics: ISSN 1361 - 1837
Price Dispersion and the Ability to Search:
The UK Internet Motor Insurance Market

Stephen McDonald and Colin Wren

Economics, Business School, 5th Floor, Ridley Building,
Newcastle University, England, UK, NE1 7RU.

19th October 2009

Abstract

The paper analyses the relationship between price dispersion and the ability to search using a non-sequential search model, where ability is measured by Internet usage. The model predicts a negative relationship between search ability and price dispersion, which is estimated using a dataset for the UK Internet motor insurance market. This comprises premium data for consumer types that vary in their Internet use, according to age, occupation and sex. Allowing for risk and propensity to search, the paper finds that search ability has a negative effect on price dispersion.

JEL Classification: L11, D83, D40

Key words: Price dispersion, search, Internet, motor insurance.

Corresponding author: Stephen McDonald, address as above, stephen.mcdonald@ncl.ac.uk.
1. Introduction

The ‘law of one price’ fails to hold in seemingly competitive goods markets, and this has long-intrigued Economists. Numerous models show that equilibrium price dispersion can arise from heterogeneous consumer search costs (e.g. Salop and Stiglitz, 1977; Varian, 1980), but the lack of easily-available data meant there were initially few empirical studies of price dispersion (Pratt et al., 1979; Carlson and Pescatrace, 1980). However, the arrival of the Internet has meant that research has flourished over recent years (e.g. Smith and Brynolfsson, 2001; Bayliss and Perloff, 2002; Brown and Goolsbee, 2002; and Baye et al., 2004). These studies find evidence of price dispersion that is consistent with a search explanation, although a direct link between individual search behaviour and price dispersion has yet to be firmly established.¹

In this paper a model of non-sequential consumer search is developed and data are taken for the Internet motor insurance market to test directly whether those consumers with characteristics associated with greater search ability face less price dispersion. A lower search cost is implied by a higher ability to search, which is measured by Internet usage. This is reasonable as more frequent users will typically be more Internet savvy and have a greater ability to search. This ‘learning-by-doing’ argument has antecedents that date back to Arrow (1962), and underpins the endogenous growth literature. A desirable feature of the motor insurance market is that the price data (i.e. premium quotes) are matched directly to the individual consumer characteristics. This means that by controlling for other factors, such as the insurance risk and the propensity to search, it is possible to establish a direct link between price dispersion and search ability.

¹ For example, Carlson and Pescatrace (1980) look across markets and find lower price dispersion for products that are more expensive or that are bought more frequently, which they attribute to search. Dahlyby and West (1986) eliminate possible non-search explanations and attribute the residual price dispersion to search, while Bayliss and Perloff (2002) rule-out price dispersion due to quality differences. Perhaps the closest to establishing a direct link between search and price dispersion is Brown and Goolsbee (2002), who look at Internet usage over time, attributing the reduction in price dispersion to an increase in Internet access, but focusing primarily on the level of prices.
The price data used in this study were gathered for a twelve month period over 2006-07 from UK motor insurance websites, including the leading price comparison site, confused.com, and give 32,255 observations on prices and associated individual characteristics. The prices are for average market conditions (e.g. fully comprehensive, age of car, mileage, etc), and allowed to vary by the age, sex and occupation of the consumer and car type. Interest is in Internet usage rather than access, assuming that the prices posted online do not reflect offline search.\(^2\) Table 1 shows the frequency of Internet use, and reveals that usage declines with age and is greater for males than for females. This suggests that age and sex signal ability to search. Further, it is reasonable that search ability will vary by occupation, as higher-skilled jobs generally require higher levels of education and computer use, therefore implying greater search ability.

While other studies have examined the Internet and insurance markets, the novelty of this paper lies in establishing a direct link between search ability and price dispersion. Of course, an important consideration in setting a premium is the risk associated with a policy, but we develop a model that predicts a negative relationship between price dispersion and search, and which is invariant to the presence of risk. The model is estimated using 1,320 price sets for 22 individual types (by age, occupation and sex) and five car types, including terms that capture risk and the propensity to search. Overall, the paper finds the ability to search is an important determinant of price dispersion, and that price dispersion decreases as the ability to search increases, particularly by age and occupation. The next section sets out the model, Section 3 outlines the data, Section 4 contains the empirical results and finally Section 5 concludes.

---
\(^2\) Motor insurance may be purchased either online or offline via brokers, shops or over the ’phone. If firms charge the same price to a given individual through each of these channels then the proportion of Internet users may affect the price distribution, à la Brown and Goolsbee (2002). This might affect the interpretation of the results, but firms discriminate between Internet and non-Internet users (Brynjolfsson and Smith, 2000), so that it is not the case.
2. The Model

The Internet motor insurance market is best characterised by a fixed sample, non-sequential search, in which consumers discover multiple prices simultaneously. This kind of search has received some attention in the literature (e.g. Wilde and Schwartz, 1979; Janssen and Moraga-Gonzalez, 2004), although it is less common than either sequential search (e.g. Rob, 1985; Stahl, 1989) or the clearinghouse model (e.g. Baye et al., 2001; Morgan et al., 2006). However, as some leading brands do not quote on a comparison website, neither of these two entirely fits the Internet motor insurance market. As a single search may yield multiple prices it cannot be sequential search, but neither is it the clearinghouse model as a single search cannot reveal all prices in the market, so that non-sequential search fits best.

2.1 Non-Sequential Search

Formally, it is supposed that the market is characterised by a single homogeneous good, and a finite set of \( n \geq 2 \) prices that can be searched, \( p = (p_1, p_2, \ldots, p_n) \), where \( 0 < p_1 < p_2 < \ldots < p_n \). There are \( N \) homogeneous individuals \( (i = 1, 2, \ldots, N) \), such that each individual \( i \) undertakes the same fixed sample search of size \( m \), where \( 1 \leq m \leq n \). When \( m = 1 \) then each price has an equal probability of being selected and a price is randomly chosen, but otherwise the search intensity increases with \( m \). Search reflects the ability of individuals.

For prices \( p \) it is useful to consider the probability distribution of searched prices as \( m \) varies, from which it is possible to derive a functional form. For a fixed sample of size \( m \) from \( n \) prices

---

3 For the main UK motor insurance price comparison site, confused.com, at the time of data collection, four leading brands did not quote on this site, comprising Esure, Churchill, Direct Line and Tesco. The last three are brands or underwritten by RBS, and together with Esure they have a combined market share of around 30% in 2002.
there are \( ^nP_m \equiv n!/ (n-m)! \) samples of non-identical prices. Of these, \( ^{n-1}P_m \) do not involve \( p_1 \), from which it follows that \( p_1 \) is selected as the lowest price on a fraction:

\[
\frac{nP_m - ^{n-1}P_m}{nP_m} \equiv \frac{m}{n}. \tag{1}
\]

Likewise, \( ^{n-1}P_m \) samples of size \( m \) do not include price \( p_1 \), of which \( ^{n-2}P_m \) samples also do not include \( p_2 \), so that \( p_2 \) is selected on a fraction:\(^4\)

\[
\frac{^{n-1}P_m - ^{n-2}P_m}{nP_m} \equiv \frac{m(n-m)}{n(n-1)}. \tag{2}
\]

Continuing in the same way, it can be shown that for a fixed sample size of \( m \) from \( n \) prices \( (m \leq n) \), the probability that price \( p_r \) \( (r = 1, 2, \ldots, n) \) is selected is:

\[
\frac{^{n-r+1}P_m - ^{n-r}P_m}{nP_m} \equiv \frac{m(n-m)(n-m-1)...(n-m-r+2)}{n(n-1)...(n-r+1)}. \tag{3}
\]

This holds for \( r \leq n-m+1 \), while if \( r > n-m+1 \) then prices \( p_{n-m+2} \) and above will never be searched as the lowest price, so that these have a zero probability.\(^5\) When \( r = 2 \) then (3) gives (2), and when \( r = 1 \) then \( r \leq n-m+1 \) holds with equality, such that (3) reduces to (1). When \( m = n \) then (3) holds for \( r \leq 1 \) only, and \( p_1 \) is always searched as the lowest price by (1).

For illustrative purposes, the probability density function of the lowest searched price is sketched in Figure 1 using (3). This supposes \( n = 7 \) and \( 1 \leq m \leq 7 \). Figure 1 shows that \( p_1 \) is always the modal price, but while the distribution is uniform when \( m = 1 \), it is more concentrated about \( p_1 \) as \( m \) increases. This means the mean price falls with search, such that it is to equal to \( p_1 \) when \( m = \)

\(^4\) When \( m = n \), so that all prices are searched, \( p_2 \) is never selected as the lowest price and this probability is zero.

\(^5\) In fact, (3) is zero when \( r = n-m+2 \) (\( m \neq 1 \)), but not defined for higher values of \( r \).
7, while the variance of the lowest searched price is also inversely related to search and tends to zero as \( m \) increases. Of interest, Figure 1 suggests that the distribution of the searched prices is remarkably similar to the exponential function, and algebraically it is represented by:

\[
f(p; m, p_1) = m \exp\left\{m(p_1 - p)\right\}
\]

(4)

where \( p_1 \) is the lowest offered price and \( p \geq p_1 \) is now continuously defined. Search increases with \( m \), where \( 1 / m \) is the scale parameter. In Appendix A it is shown \( \int_{p=p_1}^{p_\infty} f(p; m, p_1) \, dp = 1 \) and from the moment-generating function that the mean and variance of \( f \) are:\(^6\)

\[
\text{mean } f(p; m, p_1) = p_1 + \frac{1}{m} \quad \text{and} \quad \text{var } f(p; m, p_1) = \frac{1}{m^2}.
\]

(5)

This suggests (4) models the non-sequential search process very well. In particular, as the search intensity increases, the mean price falls towards \( p_1 \), while the variance tends to zero, such that in the limit \( p_1 \) is always selected as the lowest price and the mean price is \( p_1 \).

### 2.2 The Modelling Framework

The above function captures search behaviour, but the resulting distribution of searched prices (like that sketched in Figure 1) does not closely match the observed distribution of prices, where relatively few firms quote at the lower end. This is because no account has been taken of the supply side. Low-intensity search yields high searched-prices, but low demand at these prices may mean they are unprofitable and so not offered, while similarly low prices may also not be offered. To model this it is supposed that there are \( j = 1, 2, \ldots, J \) individual types and \( k = 1, 2, \ldots, K \) car types, where each sub-market \((j, k)\) is homogeneous. The car choice is exogenous,

---

\(^6\)The usual expression for the exponential function is obtained by setting \( p_1 = 0 \), in (4), in which case the mean and variance are \( 1 / m \) and \( 1 / m^2 \) (see Dolton et al, 1989).
so that consumers search non-sequentially in a single sub-market only with a search intensity $m_{jk}$.

Each sub-market has $N_{jk}$ individuals and a distribution of prices $p_{rjk}$, where $r = 1, 2, \ldots, n_{jk}$.

To model the motor insurance market we focus on a single sub-market $(j, k)$. Typically, about 24 firms quote in a UK motor insurance sub-market (see below), which on the supply side suggests a competitive market structure in which there is free entry and exit. The firms are better informed than consumers, who search, and since in practice firms offer insurance repeatedly and are able to learn the nature of their market, it is assumed that they have full information. There is a single period and a sufficient number of firms, $s = 1, 2, \ldots, S$, to cover the market, while each firm posts no more than a single price in each sub-market.\(^7\) In the knowledge of consumer search, the firms move first in posting their prices, and they are earn a net operating return $R_{jk}$ from $(j, k)$, which given free entry is constant across firms, although possibly differing across sub-markets.

As regards the firms’ costs, these are primarily fixed in nature, comprising the actuarial costings, the setting-up of a website, software development, advertising, premises, non-variable staffing costs and so on. It is supposed that each firm $s$ has a fixed cost of $F_s$. Given its market strategy, actuarial costings, fixed costs and net operating returns $R_{jk}$, a firm $s$ may decide to opt-out of some markets, so that if it quotes for $J = J\subseteq J$ individual and $K'\subseteq K$ car types its profits $\pi_s$ are:

$$\pi_s = \sum_{j \in J', k \in K'} R_{jk} - F_s$$

(6)

Firms may make super-normal profits overall, i.e. $\pi_s > 0$, but competitive pressures mean each firm makes $R_{jk}$ in each sub-market in which it quotes, so that it is unable to cross-subsidise its

\(^7\) For the purpose of exposition, for the time being we revert to the discrete version of the price set $p$.

\(^8\) If individuals tend to search on firms, then a firm’s lowest price is likely to be discovered. In practice, a firm may quote a different price on a comparison than on its own website, but below we consider a single leading comparison site (and the websites of firms not quoting on this site), as we have insufficient resources to sample all such sites. A firm may chose not to enter a sub-market, which is discussed below.
activities across sub-markets. The equilibrium occurs where, in the knowledge of consumer search, firms post their prices at no more than a single price in each sub-market \((j, k)\), but only offer that number of contracts in each sub-market to make a net return of \(R_{jk}\), such that overall each firm that posts non-zero prices is able to make non-negative profits \(\pi_i\).

Finally, the good under consideration is an insurance contract, so that the prices are premiums and in the event of an accident an individual makes a claim on the contract. Within each sub-market all contracts carry the same expected claim, but this will affect the price dispersion, as proportionately more of the lower-priced contracts must be sold in order to make a return of \(R_{jk}\), so that as the risk increases fewer firms will be observed quoting lower prices. In considering insurance it is assumed that the firms are risk neutral, so that \(R_{jk}\) now denotes the expected net operating return from sub-market \((j, k)\), including the claim, and \(\pi_i\) is the expected profit.

### 2.3 Price Distribution with Insurance

To model the price dispersion for the case of motor insurance, for sub-market \((j, k)\), let \(q_j (0 \leq q_j < 1)\) denote the probability that an individual has an accident and let \(C_k (> 0)\) denote the claim amount in the event of an accident, where without real significance these are independent of one another. At a price or premium of \(p_{rjk}\) the expected revenue on a single contract is \(p_{rjk} - q_j C_k (> 0)\), and given the above assumptions the number of firms quoting at this price is:

\[
X_{rjk} = \frac{(p_{rjk} - q_j C_k) f(p_{rjk} - q_j C_k; m_{jk}, p_{j,k}) N_{jk}}{R_{jk}}. \tag{7}
\]

\(^9\) Some cars attract higher risk drivers, but to the extent that this reflects average risk across individual types it is captured by \(C\). Likewise, some individual types \(j\) may have more severe accidents, but this is also reflected in \(C\). These are without significance as \(q\) and \(C\) are always considered together, as a product. Since insurance is a legal requirement and the level of coverage is kept constant the risk preferences of the individuals are not considered. Potentially, risk aversion could affect the propensity to search, but this is taken up below.
The price \( p_{rj} \) supports at least one firm if and only if \( X_{rj} \geq 1 \). This can be seen by noting that the first term in the numerator of (7) is the expected revenue on a single contract at price \( p_{rj} \), the second term is the probability that \( p_{rj} \) is searched as the lowest price and the final term is the market size \( N_{jk} \).\(^{10}\) The numerator therefore gives the expected total revenue that can be earned on contracts at a price \( p_{rj} \), and the denominator is return which can be earned by a single firm. For values of \( X_{rj} \) above unity more firms enter and supply the market at price \( p_{rj} \), so that letting \( p_{rj} \) vary then (7) maps out the entire distribution of prices for \((j, k)\).

This situation is sketched in Figure 2 for a level of search \( m \), letting the price vary continuously and dropping subscripts. The SS schedule plots the relationship between aggregate demand and price as a result of search, \( Q = f(p - qC : m, p_1)N \), and the BB schedule is the break-even line, showing the levels of demand above which firms enter the market at each price. Substituting the expression for \( Q \) into (7), setting \( X = 1 \) and rearranging gives \((p - qC)Q = R\), so that the BB is in fact a rectangular hyperbola. The intersection of the SS and BB schedules in Figure 2 shows the minimum and maximum prices, \( p_{min} \) and \( p_{max} \), and hence the price range. Between these prices the vertical distance between the SS and BB schedules gives the number of firms at each price. This price distribution is plausible, as now both the high and low prices are unattractive.

2.4 The Relationship between Price Dispersion and Search

As search \( m \) increases, the SS schedule in Figure 2 shifts in the manner indicated by Figure 1, causing the minimum and maximum prices to fall and the mean price to fall also. However, the effect on the variance of the prices cannot be discerned from Figure 2, so that it is examined algebraically. Substituting for \( f(p - qC : m, p_1) \) in (7) from (4), with \( p \) replaced by \( p - qC \), the distribution of prices for each sub-market is given by \((p \geq p_1 > qC)\):

\(^{10}\) To make search comparable across sub-markets \( q_jC_k \) is deducted from the premium in \( f \).
First of all, consider the standard good case, i.e. \( q = 0 \). Then, in Appendix B it is shown that the mean of (8) is inversely related to search, confirming the diagrammatic analysis. The variance is more difficult to sign, but setting the minimum price arbitrarily small, i.e. \( p_1 = 0 \), then:

\[
\frac{\partial \text{var} h(p:m, 0, 0)}{\partial m} < [>] 0 \quad \text{iff} \quad m > [<] \frac{8}{9} \frac{N}{R}.
\]  

As the search sample size \( m \) increases from unity, price dispersion at first increases but then falls. As intuition, when \( m = 1 \), so that all consumers sample a single price only, a single high-price is offered in the market as no firm has an incentive to offer less than this.\(^{11}\) However, as consumers begin to search, it is advantageous for some firms to offer lower prices to capture these and the variance increases, but as search continues the variance must eventually fall and it continues to do so. Thus, for sufficiently high levels of search relative to the market size and net return, (9) tells us that the price dispersion decreases as the search intensity increases.

With insurance, \( q > 0 \), and again setting the minimum price arbitrarily small, the variance of \( h(p:m, q C, 0) \) is derived in Appendix C as follows:\(^{12}\)

\[
\text{var} h(p:m, q C, 0) = \text{mean} h(p:m, q C, 0) [A - \text{mean} h(p:m, q C, 0)],
\]  

where

\[
\text{mean} h(p:m, q C, 0) = \exp(m q C) \frac{(2 - m q C)}{m^2} \frac{N}{R},
\]  

\(^{11}\) This is can be thought of as the monopoly price, although insurance is a legal requirement and in principle firms could charge a single price up to where car ownership ceases to be worthwhile.\(^{12}\) When \( q = 0 \) then (10) to (12) reduce to the expressions for the standard good case derived in Appendix B.
and  \[ A = \frac{6 - 2mqC}{m(2 - mqC)}. \]  (12)

For sufficiently high levels of risk \( qC \), a negative relationship exists between the mean premium and \( qC \) in (11). This is because the low-premium contracts carry the same expected claim, and so as the risk increases relatively more of these contracts must be offered to earn a net return of \( R \), causing the mean premium to fall. However, if \( 0 < mqC < 1 \) the mean premium is positively related to risk, and since this accords with intuition and Figure 2 (as \( qC \) increases the asymptote shifts to the right in this figure), this is assumed to hold. In this case the mean premium in (11) is positive and there is negative relationship with the search intensity \( m \), which is like before.

As regards the variance, differentiating (10) with respect to \( m \) gives:

\[
\frac{\partial \text{var} h(p : m, qC, 0)}{\partial m} = \frac{\partial \text{mean} h(p)}{\partial m} \left[ A - \text{mean} h(p) \right] + \text{mean} h(p) \frac{\partial A}{\partial m}. \quad (13)
\]

To sign this, \( A \) is negatively related to \( m \) in (12), while in Appendix C it is shown that:\(^{13}\)

\[
\frac{A}{2} > \text{mean} h(p : m, qC, 0) \quad \text{if} \quad m > \frac{2\exp(2 - \sqrt{2})}{1 + \sqrt{2}} \frac{N}{R} \approx \frac{3N}{2R}. \quad (14)
\]

From this it follows that the variance of the price distribution is negatively related to search \( m \) when (14) is satisfied, but that otherwise it can be positive. The term in square brackets is approximately equal to 1.5, so that it is a stronger condition than that for standard good in (10), but it is sufficient only and it reflects the effect of risk. Of interest is that it points in the same direction as the standard good case, such that an increase in search leads a reduction in price dispersion providing search is sufficiently strong. Since \( A \) is increasing in \( qC \) it can be shown that (14) is a sufficient condition for the variance to increase with risk, but at a decreasing rate.

^{13} A is decreasing in \( m \) when \( mqC < 3 - \sqrt{3} \), which is satisfied since \( 0 < mqC < 1 \).
3. The Data

The data were collected at monthly intervals over the one-year period, February 2006 to January 2007, from confused.com, the UK’s leading motor insurance comparison website. This site was launched in 2003, and at 2007 it claimed to cover 94% of the firms in this market, when it had a two-thirds share of motor insurance policies sold via an online aggregator (Financial Times, 2007). A firm is a distinct maker of quotes, where there are three principal methods of selling motor insurance, each of which is represented on confused.com. The data were collected by taking-on assumed identities in relation to personal characteristics, car type and location. The confused.com site is advantageous in this respect, as it allows search to be made at a fine level and it easily accommodates repeated search. The confused.com website listed six companies as not quoting on its site, and data were collected directly on a similar basis from the websites of four of these, Churchill, Direct Line, Esure and Tesco, each with a large market share. It is reasonable that consumers wishing to sample further prices will go to these sites.

3.1 The Customer, Car and Policy Characteristics

In relation to personal characteristics, these were chosen to reflect different search abilities, and allowed to vary by age, occupation and sex. It is hypothesised that age lowers search ability, and four ages were chosen, 25, 40, 55 and 70-years (see Table 1), which cover a good age spread, with fewer firms quoting outside this range. Since search may vary by occupation, we chose blue and white-collar occupations of a ‘factory worker’ and ‘computer consultant’, where it is

---

14 Confused.com is part of the Admiral Group, which also competes as a direct insurance provider, but no evidence of bias was found. It may be possible to infer search behaviour from the underlying programmed software, but these are commercially highly sensitive, while direct approaches to confused.com and its parent were unanswered. Preliminary investigation extended to consider other comparison sites, such as moneysupermarket.com, while more recently other sites have been launched, such as gocompare.com and comparethemarket.com.

15 An underwriter may sell insurance direct to the consumer, through an exclusive relationship with a firm or firms (e.g. a branded product or products), or through a broker that deals with a range of underwriters.
expected that the latter is more skilled at search.\textsuperscript{16} We also include an ‘unemployed individual’, who on average is likely to be less skilled at search, and a ‘retired person’, which may embody fairly heterogeneous individuals, including the inactive and so-called ‘silver-surfers’. Finally, data were also collected for males and females, giving a total of 22 consumer types.\textsuperscript{17}

The main UK motor industry body, the Society of Motor Manufacturers and Traders, defines nine car market segments, and the car types were chosen for the leading model in five of these segments, based on total UK sales in the year 2000. These are the Ford Fiesta Encore (13.3\% of sales in the Super-mini car segment in the year 2000), the Ford Focus Zetec (17.3\% of the Lower Medium segment), the Vauxhall Vectra CD 16V (14.3\% of Upper Medium cars), the BMW 525i (12.9\% of Executive cars) and the MG MGF (8.6\% of Sports cars).\textsuperscript{18} Each model was assumed to be six years old, which is the median UK car age (SMMT, 2006). Details of the car types are given in Table 2, which shows the car values (at 2006), the total sales (at 2000) and a risk index. The BMW is the highest risk car, while the Fiesta and Focus are relatively low risk.

The other factor that could be varied is the geographical location of the car owner. Preliminary work suggests motor insurance quotes can vary substantially, even from street to street, so that it was decided to choose a single address. This is for the Gosforth postcode district (NE3 2), which is a reasonably affluent suburb of Newcastle upon Tyne, and which has a mean house price close to the national pattern by four house types (i.e. flats, terraces, semi-detached and detached). To select a street in this district, 40 postcodes were randomly sampled, and a postcode chosen with a

\textsuperscript{16} The four insurers not quoting on confused.com did not permit search at this level, so that for these ‘manual’ or ‘manufacturing and engineering’ and ‘professional’ or ‘scientific and technical’ were used instead.

\textsuperscript{17} For realism, 70-year olds are retirees only, while the retired may also be 55-years of age.

\textsuperscript{18} The other segments are the Mini and Luxury vehicles, which are relatively small markets, and the Multi-Purpose and Dual Purpose vehicles (i.e. Sports Utility vehicles), but which are similar to the Upper Medium segment.
standard deviation of quoted prices across firms closest to the mean deviation. Supplementary price quote data were collected for four other locations for a single month to test the sensitivity of the results. These are for contrasting areas in the same region – one urban (Byker) and one rural (Hexham) – and two areas in the south of England that are comparable to Gosforth in their house prices and stock; the Woodley area of Reading and the Fishponds district of Bristol.

As many as possible of the other motor insurance policy details were held constant, and selected to reflect average market conditions. These are a fully comprehensive level of cover, an annual mileage of 9,000 miles and business, commuting and social use. All drivers have a 5-years no-claim bonus. Some other features were set to zero, including no-claims bonus protection and legal fees cover. Few firms include the latter as standard, although many include a courtesy car, which was therefore requested. Consumers may vary the voluntary excess amount, which is the owner’s liability in the event of a claim, but this was set to zero. Policies may also include a compulsory excess, and the data were collected for the lowest quoted premium and associated excess. Policies may vary in other details that cannot be controlled by the consumer, such as windscreen or audio equipment cover, but these are generally minor and confined to the small print, so that for practical purposes the policies are taken to be homogeneous.

3.2 Data Description

Data were gathered for 22 individual and 5 car types over a twelve-month period, giving a total of 1,320 price sets. Each price set comprises data on price quotes collected from confused.com and four directly-searched insurers, giving a total of 32,255 observations on prices. The mean number of firms quoting in a price set is 24.4, and there is little variation in this by individual or

---

19 This is Kirkley Close (NE3 2LJ). The procedure reflected the interest in price dispersion. Twenty-five firms offered prices for some or all of the 40 sampled postcodes, of which 13 offered the same price at each location. Of the remainder, the greatest standard deviation across 40 locations for any firm was 14.7% of the mean premium.

20 Prior to July 2006 confused.com sometimes reported multiple quotes for a single firm with different premium and compulsory excess combinations, but after then it quoted the lowest premium only.
The total number of firms in the dataset is 41, but entry and exit means the number quoting in each month is between 31 and 36, of which 27 quote in every month.

The mean and coefficient of variation of the premiums for each individual and car type is shown in Table 3. These are calculated across price sets for the twelve months. As would be expected, the mean premium increases with the risk index of the car types shown in Table 2, but given this there is a consistent pattern for the mean premium by the individual type. The mean premium decreases with age (except that 70-year olds pay a higher premium than 55-year olds), decreases with the occupation as we move down the table, and is higher for males. These reflect individual risk characteristics, so that higher premiums are generally paid by those that either have higher accident rates or more severe accidents, e.g. a 25 year-old male. The individual and car types may also signal information about income and the propensity to search, and we take-up this point below. As regards the coefficient of variation, several patterns are evident that appear to reflect both risk and search factors, and it is the latter that forms the basis for our empirical work.

Table 4 gives information on the compulsory excess, and shows that an excess is required in at least 80% of cases, being virtually always requested for the BMW and MGF cars. The excess amount is greatest for these two cars, although relative to the premium it is smaller than that for the Fiesta or Focus. Compared to these last two, an excess is more likely to be required for the Vectra, reflecting its risk status in Table 2, although the mean excess is smaller. The excess amount was not ascertained in about 10% of cases (Table 3), so that these are predicted. This

---

21 For the car types these range between 23.6 and 26.0, although 19.7 for the MGF. The minimum number of firms is between 11 and 19, and 30 and 31 for the maximum, although 11 and 25 firms respectively for the MGF.

22 The car type also signals information about the kinds of individual that drive that car, e.g. MGF drivers are likely to be riskier prospects than those driving family vehicles. To the extent that this does not vary according to the individual characteristics of age, occupation or sex, then it is reflected in the premium for the car type.

23 In the context of the model, a higher propensity to search means that a larger fixed sample $m$ is chosen.

24 The excess amount could be ascertained by following the web-link through to the insurer’s own website, but resource considerations prohibited this, so that these are estimated. The subsequent regressions were run both with and without the cases involving the predicted excesses, but making no qualitative difference to the results.
is based on the following regression of the excess amount ($EXCESS$) on the premium ($PREM$) with car and firm dummies (there is little variation in the excess amount by individual type).

$$EXCESS = 227.1 - 0.06 \text{PREM} - 4.5 \text{FOCUS} + 25.0 \text{VECTRA} + 102.8 \text{BMW} + 87.1 \text{MGF}$$ (15)

\[ (141.1) \quad (17.7) \quad (5.1) \quad (26.1) \quad (82.3) \quad (65.9) \]

$R^2 = 0.68$, $n = 29,055$, robust standard errors, t-ratios in parentheses, includes firm dummies.

4. Empirical Analysis

The model is estimated using the data for the 1,320 price sets, where price dispersion for each price set is measured by the coefficient of variation of the premiums in that price set ($CVPREM$).

The basic relationship is linearised, so that the estimating equation is of the form:

$$CVPREM_{jk} = \alpha - \beta_1 m_{jk} + \beta_2 N_{jk} - \beta_3 R_{jk} + \beta_4 RISK_{jk} + \text{Month dummies}_t + \varepsilon_{jk}$$ (16)

where $m$, $N$ and $R$ measure the ability to search, market size and net return respectively. To these we add variables to capture risk, monthly dummies and an error term $\varepsilon$. It is estimated with monthly data for the sub-markets ($j$, $k$), i.e. 22 individual types and 5 car types. Natural logs are taken of the premium and excess data. The variables used to measure each of the $m$, $N$, $R$ and $RISK$ are now considered, where we expect the $\beta$s in (16) to be positively signed.

4.1 The Variables

To measure the ability to search $m$, we include binary variables for the individual characteristics, age ($AGE$), occupation ($OCC$) and sex ($SEX$). As well as ability, consumers may also differ in their propensity to search, which depends on economic factors. Table 2 shows that the cars have different market values and are likely to be associated with different income groups, while Table 3 shows that individual types face different premiums. Search ability reflects innate personal
characteristics and is independent of the car type, so to pick-up differences in search propensity dummy variables are included for the car type (CAR) and for the car type interacted with the individual characteristics. The interaction terms were almost uniformly insignificant, and so are not included or reported below. This suggests that the CAR dummies alone are able to pick-up the differences in the propensity to search between sub-markets.\textsuperscript{25}

Estimation also requires knowledge of the market size $N$, which is shown in Table 2. However, this cannot be included along with the CAR dummies, so that for this reason the regressions are separately run for each car type, while recognising that when it is estimated across all types the CAR terms may also capture the effect of market size. We have no data for the size of each sub-market, while it is not possible to separately dummy this out. However, the above interaction terms between the car types and individual characteristics suggest that this is not an issue.

As regards the net return $R$, it is reasonable that this is the constant across sub-markets due to competitive pressures, but given this there may be differences in the price dispersion between price sets due to the behaviour of particular firms. First, if a firm quotes in many sub-markets then its break-even position within any one market will be lower, causing more firms to quote at that price (i.e. the line $BB$ shifts upwards in Figure 2). To capture this, terms are included for the total number of firms opting-out of each price set according to individual ($OPTOUTI$) and car type ($OPTOUTC$). These cannot be signed \textit{a priori} as the effect on the variance depends on where in the price distribution the firms offer their prices. Second, differences in the firms’ fixed costs in (6) may have similar effect on the break-even position, and so firm dummies ($FIRM$) are included for this. Third, as the number of firms ($NUMBER$) increases in a price set then the

\textsuperscript{25} The firms do not observe information on individual incomes, and it may be that the car type signals differences in the propensity to search, while the differences in individual characteristics pick-up the differences in risk, for which $RISK$ variables are included below. The interaction terms between the car types and individual characteristics were significant for the MGF car, but they pick-up the effect of the car opt-out that is included below.
extremes of the price distribution may be made even more relatively unattractive, which suggests a negative effect on price dispersion.

Finally, the model predicts that price dispersion increases at a decreasing rate with risk, i.e. $q C$ in (10). Since riskier sub-markets are characterised by higher premiums (Table 3), we measure $RISK$ using the mean premium ($MPREM$), including a quadratic term for this. Since the firm’s risk decreases as the excess ($EXCESS$) increases, then this term is also included.

4.2 The Results

The results from estimating (16) using OLS are reported in Table 5, both overall and for the five car types. The base case is a 25-year old, unemployed male, who in the equation for all car types drives a Ford Fiesta car. The model provides a good fit to the data, especially for some car types, and in the case of all car types virtually all the coefficients are significant at the 1% level. Before discussing the ability to search terms, we consider the other estimates in this table.

4.2.1 The Control terms ($n, R$ and $RISK$)

The estimates on the $CAR$ terms suggest that the car types with higher premiums face less price dispersion, and this is consistent with the propensity to search explanation. Of course, the $CAR$ terms may also capture the effect of market size, and the estimates could also be interpreted as supporting this, although the propensity to search explanation seems more plausible. If in place of the $CAR$ terms a single variable is included for the market size (according to the sales data in Table 2) then this is significant and correctly signed. As regards the firm dummies, up to 8 of the 40 these terms is significant (not reported), while in the regression for all car types the opt-outs are also significant. These suggest when firms drop-out of a car market ($OPTOUTC$) the remaining firms are more likely to quote at the extremes of the price distribution, suggesting they
are specialised in particular models, but when they drop-out from quoting for specific individuals (OPTOUTI) they tend do so for premiums at the extremes. Finally, as expected, the more firms that quote (NUMBER) in a sub-market the smaller is the price dispersion.

The coefficients on the mean premium (MPREM) terms suggest that price dispersion increases with the risk of each car type, albeit at a decreasing rate, which is significant for four out of the five cars. When estimated across all car types in the first column of Table 5 it was necessary to include slope dummy variables on the mean premium for each car type. The coefficients on the combined MPREM terms suggest that the relationship between price dispersion and risk shifts upwards for more risky cars, and it accords well with the pattern shown in Table 2. Finally, the coefficient on EXCESS is generally negative and significant, again as expected.

4.2.2 The Ability to Search terms (m)

The AGE, OCC and SEX terms capture the ability to search. In the case of age, the estimate for all cars suggests that price dispersion increases with age at the 1% level, both between the 25-year and 40-year olds, and between the 40-year and 55-year olds, suggesting that search ability has a strong effect on price dispersion. There is no significant different in dispersion between the 55-year and 70 year olds, but it could be that the number of 70-year olds using the Internet to purchase motor insurance is relatively small, so that for practical purposes the firms treat them the same as 55-year olds. If each car type is considered in Table 5 then there is a remarkably similar pattern in the relative magnitude of the coefficients for the 40, 55 and 70-year olds, with the 40-year olds always facing less price dispersion. In the case of the 25-year olds the pattern is maintained for the Fiesta and Focus cars, but there is much greater price dispersion for the other three car types. These are the higher-risk cars shown Table 2, and it seems that we are unable to

---

26 Evaluating the terms on MPREM and MPREM² using the values for the mean premium in Table 3 suggests that the overall effect is positive in each case, except for the MGF where a negative value is found. However, the model suggests a negative relationship is possible for sufficiently high levels of risk.
control for the greater risk associated with the 25-year olds driving these cars, which is despite a
5-years no-claim bonus being stipulated. Nevertheless, overall, these results suggest that price
dispersion is positively related to age, and so inversely related to ability.

For the occupation variables, the results for all cars show that price dispersion is significantly
greater for the unemployed than for all other occupations, while the retired have significantly
greater price dispersion than for the factory worker and computer consultant, each at the 1%
level. The computer consultant also faces less price dispersion than the factory worker, which is
significant at the 1% level for three of the car types. Of course, the term for the unemployed
could be picking-up greater risk, but for several reasons we do not believe this to be so. Overall,
these results tend to support the hypothesis that ‘occupations’ associated with higher
abilities to search face lower price dispersion. This is an important result, as unlike age (and the
25-year olds in particular), risk is clearly much less of a consideration here.

Finally, in the case of SEX, Table 5 shows that its effect on price dispersion is dependent on the
car type. For three of the car types this variable is insignificant, but significant at the 1% level
for the Fiesta and Focus, although varying in sign. An explanation for this may be that female
drivers of these cars differ in their search behaviour, as the Fiesta car is a small functional car
whereas the Focus is a family car. It is therefore possible that the Focus will tend be driven by
‘cost-conscious mothers’, which means it is picking-up a difference in the search propensity.

Generally, there appears to be little difference in price dispersion between males and females,
although Table 1 suggests that Internet usage is also not very different between these.

27 Unfortunately, for these cars there are no more variables available to control for the additional element of risk
associated with 25 year-olds, which the MPREM terms seem unable to wholly capture.
28 Potentially, this could be part of the story for the MGF car, where the unemployed face substantially greater price
dispersion, but the pattern of estimates is broadly similar elsewhere. In addition, the minimum premium of each
price set was regressed on the consumer characteristics (with car dummies) and no significant difference was found
between occupation types, including the unemployed, whereas that for the 25-year olds was significant and positive.
4.3 Robustness of the Results

Overall, these results suggest that lower levels of price dispersion are faced by those individuals who we expect to have higher search abilities, i.e. the younger and the higher-skilled employed.

It remains to consider the robustness of the results. Broadly, the sensitivity of the results is examined in relation to three aspects, and these are the measure of price dispersion, the measure of the price and the choice of driver location. Direct estimation of (8) is also undertaken using maximum likelihood techniques. The results of these robustness tests are given in Table 6.

First, instead of the coefficient of variation, (16) was estimated with the price range between the maximum and minimum (logged) premiums of each price set taken as the dependent variable. The results in Column 1 of Table 6 can be compared with that for all car types in Table 5. While there are some minor differences (e.g. the SEX term is now insignificant and price dispersion is now the same between the Focus and Fiesta cars), they exhibit a remarkably similar pattern in the sign, significance and relative magnitude of the coefficients, suggesting that the conclusions regarding the ability to search are unchanged. Equation (16) was also re-estimated using the ‘gap’ measure of price dispersion, i.e. the difference between the two lowest prices in a price set (see Baye et al, 2004), but the results (not shown) were poor, with a low goodness of fit and many insignificant or wrongly-signed variables. Underlying the gap measure is the Bertrand paradigm, in which some consumers have full information and purchase at the lowest price, but this sits uneasily with the motor insurance market in which not all firms quote on the comparison site, while the gap measure captures only a small part of the total price dispersion.

The price data are for the lowest quoted premium, but there is a compulsory excess in about 90% of cases (Table 4), and the regression in (15) shows that there is a negative relationship between the excess and premium. As a second exercise (16) was re-estimated but with the premium data
adjusted for the excess. Broadly, this involved finding the certainty equivalent premium for each premium-excess combination, assuming consumers have a constant relative risk aversion utility function and letting the risk parameter and consumer income take plausible values (see Appendix D). Qualitatively similar results were obtained for each parameterisation, so that a representative estimation is reported in Column 2 of Table 6, but again it supports the basic findings.

Third, motor insurance quotes can vary geographically, so that to test the sensitivity of the results data were collected for the Ford Focus car for four other locations for a single month towards the middle of the sample period (see above). The regression for the Focus car in Table 5 was re-estimated but including the additional observations and intercept dummies for each location. The result in Column 3 of Table 6 again suggests little change. When slope dummies were placed on the search terms for the Reading and Bristol locations these were not significant at the 10% level, either as a group or for the three $AGE$ and $OCC$ terms considered separately. These areas were chosen on the basis of similar house prices and stock, but two other areas were chosen because of their contrasting nature, and here significant differences were found for the slope dummies at the 5% level. In the case of Hexham, which is a prosperous, rural satellite of Newcastle, the results suggest that the unemployed face less price dispersion, but this supports the Internet-use story as these are likely to be better educated. In the case of Byker, which is a deprived urban area, the difference arises because of the 25-year-olds, but again this suggests that the estimation is unable to completely control for the greater risk associated with these individuals.

Finally, the expression for the distribution of prices in (8) was estimated directly using maximum likelihood techniques. This is potentially advantageous, as it enables the model structure to be imposed on the data. Again, Appendix D sets out the estimating approach, and the results for all

\[ \text{29} \]

\[ \text{The positive coefficient on the intercept dummy for Hexham is consistent with this, and it arises because the unemployed form the base case, and a similar explanation follows for Byker below.} \]
car types is shown in Column 4 of Table 6 using the data for all 32,255 observations. In this case \( m \) increases with search, which is inversely related to the variance in (8), so that we expect larger estimated coefficients for those with greater search ability. While there is now no significant difference between the 25 and 40-year olds, the results are supportive of those obtained using the coefficient of variation, while they also offer good support for the model.

5. Conclusions

The paper analyses the relationship between price dispersion and the ability to search using a non-sequential search model, which is estimated using data for the UK Internet motor insurance market. The model shows that there is a negative relationship between search ability and price dispersion, which is invariant to insurance risk. It is argued that the ability to search depends on Internet usage, which in turn is related to individual characteristics. The paper uses data on 1,320 price sets, involving 32,255 observations on prices and consumer characteristics, from data that was gathered from motor insurance websites for 22 consumer and five car types.

The results provide evidence of a direct link between price dispersion and ability to search. We find that younger consumers, who use the Internet more frequently, typically face lower price dispersion, while the unemployed, who are expected to have weaker search skills, face significantly greater price dispersion. Individuals in the highest-skilled occupation of computer consultant also face lower price dispersion. There is generally no difference between males and females, although their Internet usage does not vary greatly. Overall, the paper finds that price dispersion is lower for those consumer types with greater Internet usage, suggesting that price dispersion is inversely related to the ability to search.
Potentially, there could be other interpretations of the results. One of these could be that greater price dispersion arises from poorer Internet access, although for this to occur online and offline prices would have to be identical, but we do not believe this to be the case. Further, greater risk could lead to greater price dispersion, but our results are not consistent with this explanation, and they are obtained for the low-risk cars where this does seem an issue. Thus, we believe we have good evidence of a negative relationship between the ability to search and price dispersion. It is plausible that this behaviour will be observed in other markets, while the policy implication is that in order to increase consumer welfare it may not be sufficient just to get people online.
Table 1: Frequency of Internet Use by Age and Sex

<table>
<thead>
<tr>
<th>Age:</th>
<th>Daily</th>
<th>Weekly</th>
<th>Less frequently</th>
</tr>
</thead>
<tbody>
<tr>
<td>16-24 years</td>
<td>77</td>
<td>19</td>
<td>4</td>
</tr>
<tr>
<td>25-44</td>
<td>72</td>
<td>20</td>
<td>8</td>
</tr>
<tr>
<td>45-64</td>
<td>65</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>65 plus</td>
<td>54</td>
<td>27</td>
<td>19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sex:</th>
<th>Daily</th>
<th>Weekly</th>
<th>Less frequently</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>73</td>
<td>20</td>
<td>7</td>
</tr>
<tr>
<td>Female</td>
<td>65</td>
<td>24</td>
<td>11</td>
</tr>
</tbody>
</table>

Notes: Internet use within previous three months. Usage data not available by occupation.
Table 2: Car Types

<table>
<thead>
<tr>
<th>Make and model</th>
<th>Engine size (cc)</th>
<th>Car value</th>
<th>Total sales</th>
<th>Risk rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ford Fiesta Encore</td>
<td>1299</td>
<td>£1,595</td>
<td>91,783</td>
<td>4</td>
</tr>
<tr>
<td>Ford Focus Zetec</td>
<td>1596</td>
<td>£3,350</td>
<td>114,512</td>
<td>5</td>
</tr>
<tr>
<td>Vauxhall Vectra CD 16V</td>
<td>1998</td>
<td>£3,120</td>
<td>70,704</td>
<td>12</td>
</tr>
<tr>
<td>BMW 525i</td>
<td>2494</td>
<td>£8,065</td>
<td>13,443</td>
<td>16</td>
</tr>
<tr>
<td>MG MGF</td>
<td>1796</td>
<td>£4,320</td>
<td>5,766</td>
<td>12</td>
</tr>
</tbody>
</table>

Notes. 1. parkers.co.uk, 6th February 2006. 2. Sales at 2000, SMMT (2006). 3. Advisory risk rating index of the Association of British Insurers, based on damage and parts costs, repair times, new car values, body shells, performance and security. The index lies between 1 and 20, where 1 is the lowest risk.
## Table 3: Descriptive Statistics for Premium by Car and Individual Type

<table>
<thead>
<tr>
<th>Car type:</th>
<th>Fiesta</th>
<th>Focus</th>
<th>Vectra</th>
<th>BMW</th>
<th>MGF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 25 years</td>
<td>373</td>
<td>389</td>
<td>570</td>
<td>905</td>
<td>692</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.12)</td>
<td>(0.14)</td>
<td>(0.22)</td>
<td>(0.25)</td>
</tr>
<tr>
<td>40 years</td>
<td>269</td>
<td>281</td>
<td>403</td>
<td>526</td>
<td>465</td>
</tr>
<tr>
<td></td>
<td>(0.13)</td>
<td>(0.13)</td>
<td>(0.12)</td>
<td>(0.11)</td>
<td>(0.22)</td>
</tr>
<tr>
<td>55 years</td>
<td>212</td>
<td>223</td>
<td>306</td>
<td>400</td>
<td>371</td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
<td>(0.16)</td>
<td>(0.16)</td>
<td>(0.13)</td>
<td>(0.28)</td>
</tr>
<tr>
<td>70 years</td>
<td>223</td>
<td>232</td>
<td>325</td>
<td>414</td>
<td>384</td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
<td>(0.18)</td>
<td>(0.18)</td>
<td>(0.16)</td>
<td>(0.26)</td>
</tr>
<tr>
<td>Occupation Unemployed</td>
<td>299</td>
<td>316</td>
<td>455</td>
<td>653</td>
<td>577</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.13)</td>
<td>(0.14)</td>
<td>(0.16)</td>
<td>(0.42)</td>
</tr>
<tr>
<td>Factory</td>
<td>281</td>
<td>293</td>
<td>417</td>
<td>600</td>
<td>486</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.14)</td>
<td>(0.15)</td>
<td>(0.16)</td>
<td>(0.17)</td>
</tr>
<tr>
<td>Consultant</td>
<td>276</td>
<td>286</td>
<td>410</td>
<td>581</td>
<td>473</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.13)</td>
<td>(0.14)</td>
<td>(0.15)</td>
<td>(0.17)</td>
</tr>
<tr>
<td>Retired</td>
<td>215</td>
<td>225</td>
<td>312</td>
<td>402</td>
<td>371</td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
<td>(0.18)</td>
<td>(0.18)</td>
<td>(0.15)</td>
<td>(0.24)</td>
</tr>
<tr>
<td>Sex Male</td>
<td>278</td>
<td>290</td>
<td>412</td>
<td>584</td>
<td>495</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.14)</td>
<td>(0.14)</td>
<td>(0.15)</td>
<td>(0.25)</td>
</tr>
<tr>
<td>Female</td>
<td>267</td>
<td>280</td>
<td>400</td>
<td>562</td>
<td>448</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.14)</td>
<td>(0.15)</td>
<td>(0.16)</td>
<td>(0.25)</td>
</tr>
<tr>
<td>All individuals</td>
<td>273</td>
<td>285</td>
<td>406</td>
<td>573</td>
<td>486</td>
</tr>
<tr>
<td></td>
<td>(0.15)</td>
<td>(0.14)</td>
<td>(0.15)</td>
<td>(0.15)</td>
<td>(0.25)</td>
</tr>
</tbody>
</table>

Notes. Mean premiums in £’s, rounded to nearest £, and coefficient of variation in parentheses. Each of these calculated as the mean across the relevant price sets. Excludes compulsory excesses.
Table 4: Descriptive Statistics for Compulsory Excess

<table>
<thead>
<tr>
<th>Car type</th>
<th>Excess required (%)</th>
<th>Mean excess amount (£)</th>
<th>CV of excess amount</th>
<th>Ratio of excess to premium</th>
<th>Amount not known (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fiesta</td>
<td>80.9</td>
<td>159.0</td>
<td>0.47</td>
<td>0.58</td>
<td>13.4</td>
</tr>
<tr>
<td>Focus</td>
<td>81.3</td>
<td>152.8</td>
<td>0.46</td>
<td>0.54</td>
<td>15.7</td>
</tr>
<tr>
<td>Vectra</td>
<td>95.6</td>
<td>141.9</td>
<td>0.56</td>
<td>0.35</td>
<td>10.8</td>
</tr>
<tr>
<td>BMW</td>
<td>98.9</td>
<td>202.6</td>
<td>0.47</td>
<td>0.35</td>
<td>9.9</td>
</tr>
<tr>
<td>MGF</td>
<td>98.9</td>
<td>182.8</td>
<td>0.47</td>
<td>0.38</td>
<td>5.1</td>
</tr>
</tbody>
</table>

Notes. Final column shows % of cases in first column where excess amount is not known.
Table 5: Regression Results for Coefficient of Variation

<table>
<thead>
<tr>
<th>Dep. vble: CVPREM</th>
<th>All cars</th>
<th>Fiesta</th>
<th>Focus</th>
<th>Vectra</th>
<th>BMW</th>
<th>MGF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-704.19***</td>
<td>-87.56</td>
<td>-920.27***</td>
<td>-628.18***</td>
<td>-832.66***</td>
<td>-461.95***</td>
</tr>
<tr>
<td>AGE: 40</td>
<td>11.94***</td>
<td>8.51***</td>
<td>9.80***</td>
<td>-0.60</td>
<td>-8.07***</td>
<td>-11.62***</td>
</tr>
<tr>
<td>AGE: 55</td>
<td>25.52***</td>
<td>14.91***</td>
<td>20.49***</td>
<td>5.77*</td>
<td>0.19</td>
<td>-7.66**</td>
</tr>
<tr>
<td>AGE: 70</td>
<td>25.74***</td>
<td>13.75***</td>
<td>18.95***</td>
<td>5.33</td>
<td>1.29</td>
<td>-3.55</td>
</tr>
<tr>
<td>OCC: Factory</td>
<td>-12.35***</td>
<td>-0.65</td>
<td>-4.51**</td>
<td>-3.51*</td>
<td>-5.90**</td>
<td>-21.03***</td>
</tr>
<tr>
<td>OCC: Computer</td>
<td>-12.66***</td>
<td>-1.43</td>
<td>-4.51**</td>
<td>-3.74**</td>
<td>-7.09***</td>
<td>-22.14***</td>
</tr>
<tr>
<td>OCC: Retired</td>
<td>-10.17***</td>
<td>1.59</td>
<td>-2.60</td>
<td>-0.92</td>
<td>-5.55**</td>
<td>-18.29***</td>
</tr>
<tr>
<td>SEX: Female</td>
<td>-1.06*</td>
<td>2.73***</td>
<td>-1.46***</td>
<td>-0.32</td>
<td>-0.21</td>
<td>-0.57</td>
</tr>
<tr>
<td>CAR: Focus</td>
<td>-27.01***</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CAR: Vectra</td>
<td>-154.86***</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CAR: BMW</td>
<td>-280.15***</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CAR: MGF</td>
<td>-237.17***</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>OPTOUTC</td>
<td>2.51***</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>OPTOUTI</td>
<td>-2.85***</td>
<td>-0.11</td>
<td>-1.48**</td>
<td>-0.53</td>
<td>-1.71**</td>
<td>0.18</td>
</tr>
<tr>
<td>NUMBER</td>
<td>-0.51***</td>
<td>-0.18*</td>
<td>-0.28**</td>
<td>-0.30**</td>
<td>-0.17</td>
<td>-0.90***</td>
</tr>
<tr>
<td>MPREM</td>
<td>254.49***</td>
<td>27.71</td>
<td>322.39***</td>
<td>215.35***</td>
<td>278.74***</td>
<td>174.29***</td>
</tr>
<tr>
<td>MPREM²</td>
<td>-20.95***</td>
<td>-1.20</td>
<td>-26.62***</td>
<td>-17.39***</td>
<td>-21.27***</td>
<td>-15.19***</td>
</tr>
<tr>
<td>MPREM * Focus</td>
<td>4.29***</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MPREM * Vectra</td>
<td>24.74***</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MPREM * BMW5</td>
<td>43.30***</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MPREM * MGF</td>
<td>34.97***</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>EXCESS</td>
<td>-5.85***</td>
<td>-2.10</td>
<td>-4.21**</td>
<td>-1.37</td>
<td>-7.54***</td>
<td>4.48*</td>
</tr>
<tr>
<td>FIRM dummies</td>
<td>included</td>
<td>included</td>
<td>included</td>
<td>included</td>
<td>included</td>
<td>included</td>
</tr>
<tr>
<td>MONTH dummies</td>
<td>included</td>
<td>included</td>
<td>included</td>
<td>included</td>
<td>included</td>
<td>included</td>
</tr>
<tr>
<td>F</td>
<td>106.25</td>
<td>332.03***</td>
<td>137.96***</td>
<td>47.55***</td>
<td>87.06</td>
<td>294.49***</td>
</tr>
<tr>
<td>R²</td>
<td>0.726</td>
<td>0.920</td>
<td>0.909</td>
<td>0.765</td>
<td>0.865</td>
<td>0.954</td>
</tr>
<tr>
<td>Observations</td>
<td>1,313</td>
<td>263</td>
<td>263</td>
<td>263</td>
<td>262</td>
<td>262</td>
</tr>
</tbody>
</table>

Notes: OLS regression with robust standard errors across price sets. * significant at 10% level. ** = 5% and *** = 1% level. Natural logarithms are taken of the premium and excess.
Table 6: Robustness of the Results

<table>
<thead>
<tr>
<th></th>
<th>Column 1</th>
<th>Column 2</th>
<th>Column 3</th>
<th>Column 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-479.39***</td>
<td>-1,089.08***</td>
<td>-166.80***</td>
<td>94.97***</td>
</tr>
<tr>
<td>(AGE: 40)</td>
<td>8.36***</td>
<td>14.25***</td>
<td>9.72**</td>
<td>-2.07</td>
</tr>
<tr>
<td>(AGE: 55)</td>
<td>17.69***</td>
<td>31.67***</td>
<td>18.89***</td>
<td>-10.97***</td>
</tr>
<tr>
<td>(AGE: 70)</td>
<td>17.53***</td>
<td>31.23***</td>
<td>20.16***</td>
<td>-9.81***</td>
</tr>
<tr>
<td>(OCC: Factory)</td>
<td>-7.85***</td>
<td>-10.20***</td>
<td>-5.10</td>
<td>6.11***</td>
</tr>
<tr>
<td>(OCC: Computer)</td>
<td>-7.95***</td>
<td>-10.35***</td>
<td>-5.18</td>
<td>6.80***</td>
</tr>
<tr>
<td>(OCC: Retired)</td>
<td>-6.86***</td>
<td>-7.58***</td>
<td>-4.41</td>
<td>5.88***</td>
</tr>
<tr>
<td>(SEX: Female)</td>
<td>0.40</td>
<td>-1.46***</td>
<td>-2.13*</td>
<td>-0.73</td>
</tr>
<tr>
<td>(CAR: Focus)</td>
<td>-4.12</td>
<td>-32.31***</td>
<td>-</td>
<td>2.26*</td>
</tr>
<tr>
<td>(CAR: Vectra)</td>
<td>-81.05***</td>
<td>-171.82***</td>
<td>-</td>
<td>0.54</td>
</tr>
<tr>
<td>(CAR: BMW)</td>
<td>-118.90***</td>
<td>-329.96***</td>
<td>-</td>
<td>7.17***</td>
</tr>
<tr>
<td>(CAR: MGF)</td>
<td>-142.70***</td>
<td>-267.71***</td>
<td>-</td>
<td>60.73**</td>
</tr>
<tr>
<td>(OPTOUTC)</td>
<td>1.99***</td>
<td>1.99**</td>
<td>-</td>
<td>-7.87***</td>
</tr>
<tr>
<td>(OPTOUTI)</td>
<td>-1.91***</td>
<td>-2.43***</td>
<td>-2.42**</td>
<td>-</td>
</tr>
<tr>
<td>(NUMBER)</td>
<td>-0.14***</td>
<td>-0.56***</td>
<td>0.01</td>
<td>-0.28</td>
</tr>
<tr>
<td>(MPREM)</td>
<td>164.38***</td>
<td>357.25***</td>
<td>68.58*</td>
<td>-</td>
</tr>
<tr>
<td>(MPREM^2)</td>
<td>-13.11***</td>
<td>-28.25***</td>
<td>-4.84</td>
<td>-</td>
</tr>
<tr>
<td>(MPREM * Focus)</td>
<td>0.46</td>
<td>5.26***</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(MPREM * Vectra)</td>
<td>12.46***</td>
<td>27.03***</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(MPREM * BMW5)</td>
<td>29.65***</td>
<td>49.78***</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(MPREM * MGF)</td>
<td>25.21***</td>
<td>39.24***</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(EXCESS)</td>
<td>-2.86***</td>
<td>-</td>
<td>-7.08**</td>
<td>0.03***</td>
</tr>
<tr>
<td>(LOC: Byker)</td>
<td>-</td>
<td>-</td>
<td>-2.64</td>
<td>-</td>
</tr>
<tr>
<td>(LOC: Hexham)</td>
<td>-</td>
<td>-</td>
<td>6.20**</td>
<td>-</td>
</tr>
<tr>
<td>(LOC: Reading)</td>
<td>-</td>
<td>-</td>
<td>2.03</td>
<td>-</td>
</tr>
<tr>
<td>(LOC: Bristol)</td>
<td>-</td>
<td>-</td>
<td>1.84</td>
<td>-</td>
</tr>
<tr>
<td>(FIRM) dummies</td>
<td>included</td>
<td>included</td>
<td>included</td>
<td>included</td>
</tr>
<tr>
<td>(MONTH) dummies</td>
<td>included</td>
<td>included</td>
<td>included</td>
<td>included</td>
</tr>
<tr>
<td>(F [\chi^2])</td>
<td>80.91***</td>
<td>74.85***</td>
<td>37.41***</td>
<td>[934.6***]</td>
</tr>
<tr>
<td>(R^2 [\text{logL}])</td>
<td>0.684</td>
<td>0.696</td>
<td>0.444</td>
<td>[-32,570]</td>
</tr>
<tr>
<td>Observations</td>
<td>1,313</td>
<td>1,313</td>
<td>527</td>
<td>32,255</td>
</tr>
</tbody>
</table>

Notes: Columns 1 to 3 report OLS results with robust standard errors. Column 1 uses the premium range as the dependent variable and Columns 2 and 3 use the coefficient of variation, where Column 2 adjusts the premium for the excess assuming \(\alpha = 0.5\) and \(w_e = \£20,000\) (see Appendix D), and Column 3 is regressed for the Focus car only but including an extra 66 observations for each of four locations. Column 4 regresses the model using maximum likelihood techniques, where the opposite signs are expected on \(AGE\), \(OCC\) and \(SEX\). Logs are taken of the premium and excess. * significant at 10% level. ** = 5% and *** = 1% level.
Figure 1: Probability Density Functions for Lowest Searched Price

![Graph showing probability density functions for lowest searched price.](image)

Note: Evaluation of expression (3) for $n = 7$ and $m$ varies.

Figure 2: The Observed Price Distribution

![Graph showing the observed price distribution.](image)
Appendix A: The Mean and Variance of \( f(p: m, p_1) = m \exp \{m (p_1 - p)\} \)

The moment-generating function \( M(t) \) of \( f(p: m, p_1) = m \exp \{m (p_1 - p)\} \) is given by the definite integral, where \( t < m \):

\[
M(t) = \int_{p=p_1}^{p=\infty} m \exp \{m p_1 + (t-m)p\} \, dp = \exp \left\{ m \frac{p_1 + (t-m)p}{t-m} \right\}_{p=p_1}^{p=\infty} . \tag{A1}
\]

The solution to this is:

\[
M(t) = \frac{-m \exp(t p_1)}{t-m} . \tag{A2}
\]

When \( t = 0 \), then \( M(0) = 1 \), so that \( \int f(p: m, p_1) \, dp = 1 \). Further, differentiating (A2) gives:

\[
M'(t) = \frac{m \left[ 1 - (t-m)p_1 \right] \exp(t p_1)}{(t-m)^2} , \quad \text{and}
\]

\[
M^*(t) = -m \frac{2 \left( 1 - (t-m)p_1 \right) + (t-m)^2 p_1^2}{(t-m)^3} \exp(t p_1) .
\]

Evaluating these at \( t = 0 \), the mean of \( f(p: m, p_1) \) is:

\[
M'(0) = p_1 + \frac{1}{m},
\]

and the variance of \( f(p: m, p_1) \) is:

\[
M^*(0) - M'(0)^2 = \frac{1}{m^2} .
\]
Appendix B: The Variance of $h(p; m, 0, p_1)$

From equation (8), with $q = 0$, we have:

$$h(p; m, 0, p_1) = p m \exp \left\{ m(p_i - p) \right\} \frac{N}{R}, \quad (p \geq p_i).$$

This has a moment-generating function ($t < m$):

$$M(t) = \int_{p=p_1}^{p=m} p m \exp \left\{ m p_i + (t-m)p \right\} \frac{N}{R} \, dp. \tag{B1}$$

To solve this, differentiate $(a p + b) \exp \left\{ m p_i + (t-m)p \right\} \frac{N}{R}$ with respect to $p$ to get:

$$\{a + b(t-m) + a(t-m)p \} \exp \left\{ m p_i + (t-m)p \right\} \frac{N}{R}.$$  

Comparison with (B1) suggests $a + b(t-m) = 0$ and $a(t-m)p = p_1 m$, so that:

$$M(t) = \frac{N}{R} \left[ \left\{ \frac{p m}{(t-m)} - \frac{m}{(t-m)^2} \right\} \exp \left\{ m p_i + (t-m)p \right\} \right]_{p=p_1}^{p=m}.$$  

Since $t < m$, this can be evaluated as:

$$M(t) = \exp (t p_i) \left\{ \frac{m-m(t-m)p_1}{(t-m)^2} \right\} \frac{N}{R}.$$  

The first and second derivatives are:

$$M'(t) = -m \exp (t p_i) \left\{ \frac{z-1}{(t-m)} + 1 \right\} \frac{N}{R}, \quad \text{and}$$  

$$M''(t) = -m \exp (t p_i) \left\{ \frac{z^3 - 3 z^2 + 6 z - 6}{(t-m)^2} \right\} \frac{N}{R}, \quad \text{where } z \equiv p_i (t-m).$$

This means:
mean \( h(p : m, 0, p_1) = M'(0) = \left[ \frac{(1 - z)^2 + 1}{m^2} \right] \frac{N}{R}, \) and

\[
\text{var} \ h(p : m, 0, p_1) = M''(0) - M'(0) \varphi = \left[ \frac{z^3 + 3z^2 + 6z + 6}{m^3} \right] \frac{N}{R} - \left[ \frac{z^2 - 2z + 2}{m^2} \right] \left( \frac{N^2}{R^2} \right), \quad (B3)
\]

where \( z \) is now equal to \(-m p_1 < 0\). By (B2), the mean is inversely related to search \( m \), but in the case of the variance the relationship with search is ambiguous. To simplify, set the minimum price \( p_1 \) equal to zero, i.e. \( z = 0 \), in which case (B3) and (B2) can be written as:

\[
\text{var} \ h(p : m, 0, 0) = \text{mean} \ h(p : m, 0, 0) \left[ \frac{3}{m} - \text{mean} \ h(p : m, 0, 0) \right] \quad (B4)
\]

where

\[
\text{mean} \ h(p : m, 0, 0) = \frac{2}{m^2} \frac{N}{R}. \quad (B5)
\]

Substituting (B5) into (B4) and differentiating gives:

\[
\frac{\partial \text{var} \ h(p : m, 0, 0)}{\partial m} < (>) 0 \quad \text{iff} \quad m > (<) \frac{8 N}{9 R}. \quad (B6)
\]

Since \( \text{var} \ h > 0 \), then \( \text{mean} \ h < 3 / m \) in (B4), which using (B5) gives \( m > (2 / 3) (N / R) \), and this is consistent with (B6).
Appendix C: The Model including Risk ($q > 0$)

From (8) we get ($p > q$):

$$h(p : m, qC, 0) = \frac{(p-q)C \exp\{-m(p-q)\}N}{R}.$$ 

The moment-generating function is:

$$M(t) = \int_{p=0}^{p=x} \frac{(p-q)C \exp\{(t-m)p + mqC\}N}{R} dp.$$ 

To solve this, we proceed in the same way as Appendix B, to get:

$$M(t) = m \exp(mqC) \left\{\frac{(t-m)qC + 1}{(t-m)^2}\right\} \frac{N}{R}.$$ 

The first and second derivatives are:

$$M'(t) = -m \exp(mqC) \left\{\frac{(t-m)q + 2}{(t-m)^3}\right\} \frac{N}{R} \quad \text{and} \quad M''(t) = m \exp(mqC) \left\{\frac{2(t-m)qC + 6}{(t-m)^4}\right\} \frac{N}{R}.$$ 

The mean of $h(p : m, qC, 0)$, which is equal to $M'(0)$, is therefore:

$$\text{mean} h(p : m, qC, 0) = \exp(mqC) \frac{(2-mqC)}{m^2} \frac{N}{R}. \quad (C1)$$

Further, since $M''(0) = A M'(0)$, where:

$$A = \frac{6 - 2mqC}{m(2-mqC)}, \quad (C2)$$

then the variance, which is equal to $M''(0) - M'(0)^2$, is:

$$\text{var} h(p : m, qC, 0) = \text{mean} h(p : m, qC, 0) [A - \text{mean} h(p : m, qC, 0)]. \quad (C3)$$
Differentiating the variance in (C3) with respect to the mean gives:

\[
\frac{\partial \text{var } h(p;m)}{\partial \text{mean } h(p;m)} = A - 2 \text{mean } h(p;m),
\]

and using (C1) and (C2) to substitute into the right-hand side of this, it is positive if and only if:

\[
\frac{N}{R} < \frac{m(3-mqC)}{(2-mqC)^2 \exp(mqC)}. \quad \text{(C4)}
\]

Let \( D \) denote the right-hand side of (C4) and write \( X \equiv mqC \), then:

\[
\frac{\partial D}{\partial X} = \frac{-m(2-X)^2 \exp X - m(3-X)\left\{-2(2-X)\exp X + (2-X)^2 \exp X\right\}}{(2-X)^4 \left(\exp X\right)^3}. \quad \text{(C5)}
\]

The right-hand side numerator of (C5) has roots \( X = 2 - \sqrt{2}, 2 \) and \( 2 + \sqrt{2} \), but only the first is in the admissible range, \( 0 < X < 1 \), which gives a minimum. Thus, a sufficient condition for (C4) is obtained by putting \( mqC = 2 - \sqrt{2} \) into this expression, which after rearrangement gives:

\[
m > \frac{2\exp\left(2-\sqrt{2}\right)N}{1+\sqrt{2}} \frac{1}{R}.
\]
Appendix D: Robustness Tests

(a) The Excess

The certainty equivalent premium \( ADJPREM \) for each premium-excess combination is:

\[
q u (w - ADJPREM) = q u (w - PREM - EXCESS) + (1 - q) u (w - PREM). \quad (D1)
\]

Consumers have a constant relative risk aversion utility function, \( u(w) = w^{1-\alpha} / (1-\alpha) \), where the risk parameter is \( \alpha > 0 (\alpha \neq 1) \) and \( w \) is wealth. Substituting the utility function into (D1), rearranging, and taking a second-order Taylor expansion gives:

\[
ADJPREM = PREM + q EXCESS + \frac{\alpha [q EXCESS^2 - (ADJPREM - PREM)^2]}{2 (w - PREM)}. \quad (D2)
\]

The adjusted premium is sum of the actual premium, the expected excess amount and a term that includes the risk preference parameter \( \alpha \). \( ADJPREM \) was separately calculated for males and females and those aged 25, 40, 55 and 70 years of age, using data on published accident rates. Values of \( \alpha \) were taken equal to 0.5 and 1.5, and \( w \) equal to £10,000 and £20,000, i.e. about the minimum and median wage. Equation (D2) includes \( PREM_{\text{adj}} \) on the right-hand side, but this was initially set equal to \( PREM \) and iterated, but making no difference to the obtained results.

(b) Maximum Likelihood Estimation

Writing \( \tilde{p} = p - q C \) then (8) is \( h(\tilde{p} : m, p_1) = \tilde{p} m \exp\{m (p_1 - \tilde{p})\} N / R \), and assuming there are \( n \) i.i.d. observations on prices \( \tilde{p}_1, \tilde{p}_2, ..., \tilde{p}_n \) the likelihood function is \( L(m, p_1) = \prod_{r=1}^{n} h(\tilde{p}_r : m, p_1) \), and the log-likelihood function, \( l(m, p_1) \) is:
To estimate this we write \( m = \alpha + \beta x \), where \( \alpha \) and \( \beta \) are estimation parameters, and the \( x \) are the variables determining the search ability, i.e. age (\( AGE \)), occupation (\( OCC \)) and sex (\( SEX \)). It is estimated by running a Stata do-file on a D1 maximum likelihood program (Gould et al, 2006).

Two simplifications are made. First, to adjust the prices \( \tilde{p} \) for the risk associated with each sub-market, \( q C \), the minimum premium of the price set is used, as the mean premium eliminates the variation from the data and it will not converge. We divide through by this, so that the \( p_i \) term cancels in (D3) and it is treated as a constant. Second, in order to achieve convergence the variables for \( N \) and \( R \) terms in (D3) had to added to the regression in the same manner as the \( \alpha \) term, but since the individual opt-opt was dropped due to collinearity.

\[
I(m, p_i) = \sum_{r=1}^{r=n} \ln(\tilde{p}_r m) + n m \sum_{r=1}^{r=n} \tilde{p}_r m + n \ln N - n \ln R . \tag{D3}
\]
References


