Journal of Econometrics 29 (1985) 121–148. North-Holland

LIKELIHOOD DIAGNOSTICS AND BAYESIAN ANALYSIS OF A MICRO-ECONOMIC DISEQUILIBRIUM MODEL FOR RETAIL SERVICES*

Peter KOOIMAN

Central Bureau of Statistics, The Hague, The Netherlands

Herman K. VAN DIJK

Harvard University and NBER, Cambridge, MA 02138, USA
Erasmus University Rotterdam, 3000 DR Rotterdam, The Netherlands

A. Roy THURIK

Research Institute for Small- and Medium-Sized Business, Zoetermeer, The Netherlands

In this paper we apply Maximum Likelihood and Bayesian methods to explain differences in floorspace productivity among retail establishments in the grocery trade. The model we develop is a switching model where sales are either supply-determined or demand-determined. Under excess supply the model allows for so-called 'trading-down', i.e., an increase in the share of selling area, and, thereby, a decrease in service level.

To estimate our model we employ a cross-section of observations on individual shops. We present maximum likelihood results, and also study the shape of the likelihood surface by means of Monte Carlo numerical integration methods. With a uniform prior we obtain marginal posterior density functions both of the parameters of interest and of the average probability of the excess supply regime in the sample. The average probability of excess supply is 0.23, with a standard deviation of 0.06. This shows that, according to our estimates, excess demand is the rule and excess supply the exception in the sample that we analyse.

1. Introduction

Since 1973 the Dutch Research Institute for Small- and Medium-Sized Business (Economisch Instituut voor het Midden- en Kleinbedrijf, EIM) and the Econometric Institute of the Erasmus University Rotterdam cooperate in a research project aiming at an econometric analysis of the small- and medium-sized business sector in the Netherlands. Research effort has focussed so far on

*This study is a part of the research projects 'Fundamental Research of Small- and Medium-sized Business' and 'Equilibrium and Disequilibrium in Supply and Demand', financed by the Dutch Ministry of Education and Science (Voorwaardelijke Financiering). The second author was in part supported by a NATO science fellowship, visiting Harvard University and the National Bureau of Economic Research.

The views expressed in this paper are those of the authors and do not necessarily reflect the policies of the Netherlands Central Bureau of Statistics.

The authors are grateful to Adrie S. Louter for performing the numerical integrations by Monte Carlo simulation. Guus Boender took care of the figures of marginal posterior density functions.

the explanation of firm behaviour in the retail trade. Sales level, labour volume, floorspace, floorspace partitioning, price-setting behaviour and financial structure are, or will be, the subject of analysis. Data used for this purpose are available at the individual shop level. They stem from surveys that have been held by EIM for a great number of branches in the retail industry. The present study is a part of this project.

In this paper we apply maximum likelihood and Bayesian methods to analyse an extension of a model developed by Thurik and Koerts (1984a,b) and Thurik (1984) to explain differences in floorspace productivity, measured as sales per square meter, among individual retail establishments (= shops). In this model floorspace productivity is related to a partitioning of the floorspace into selling area and remaining space used for storage, administration, own production, staff facilities, etc. Both selling area and remaining space are treated as inputs in a production technology for retail services. The model has been applied to a wide variety of Dutch and French supermarkets and supermarket-like establishments and to several Dutch non-supermarket shop types. Thurik (1984) has used the model to investigate whether French supermarkets maximize profits or sales. He also analysed the influence of environmental factors on floorspace productivity, e.g., per capita consumer spending, population density, number of competitors, and shopping centre characteristics. No outside influences could be established, however, that were both plausible and statistically significant. Only supply factors, i.e., establishment properties, seem to play a role.

A reason for this rather counter-intuitive result may be that the model used was not appropriate. Sales, hence floorspace productivity, are determined by the interplay of supply and demand, and cannot properly by analysed from the supply side alone.[1] In leaving out a demand side, as Thurik (1984) does, one implicitly assumes that demand is always large enough to sustain sales maximization constrained by technical possibilities alone. This may not always be true actually. Therefore we introduce an explicit demand side in the model. When demand is large enough, sales will be supply-determined and the former model applies. With demand too small, however, sales will be demand-determined, and we substitute another model. Thus we end up with a switching model, where sales, and the partitioning of the floorspace, are either supply-determined or demand-determined. As we do not know which of the two regimes applies to each one of the available observations, we have to include both possibilities in the model, leaving the data to decide on the most likely regime distribution. To the extent that sensible results can be obtained, our model may serve as a framework for a renewed attempt to establish the influence of environmental factors on floorspace productivity.

[1] Thurik (1984, pp. 121–123) advances several other possible explanations as well.

Switching models with endogenous regime choice have mainly been used to analyse markets in disequilibrium, where it is assumed that transactions equal the minimum of supply and demand. They have been used to analyse, for example, the labour market, the housing market, the credit market. See a.o. Rosen and Quandt (1978), Fair and Jaffee (1972), Laffont and Garcia (1977). As far as we know only *aggregate time series* data have been used to estimate these models. It has often been argued that for this type of data the discrete switch is inappropriate as in the aggregate regime switches occur only gradually and incompletely. In actual aggregate markets supply constraints and demand constraints will always coexist. Smoothed versions of the switching model have therefore been developed by Muellbauer (1978), Kooiman and Kloek (1979), Malinvaud (1982), Lambert (1984) and Kooiman (1984), aggregating over micro markets in disequilibrium. Their basic assumption is that the minimum principle is more likely to apply to the micro level of individual markets, households or firms, than to the aggregate. The alternative is, of course, to stick to the discrete minimum condition and apply the model to genuine micro data. This is the approach we take in the present paper, where we present estimates of the discrete version of our switching model based on a *cross-section* of observations pertaining to *individual* retail establishments.

The likelihood function of the discrete switching model is rather complicated, so that no analytical results can be obtained and numerical optimization has to be used to find maximum likelihood estimates. A complication arises from the fact that the likelihood function can be shown to tend to infinity for suitable values of the structural parameters, when variances tend to zero. This feature, which does not exclude a consistent root of the likelihood equations corresponding to a local maximum, may seriously frustrate the application of numerical optimization techniques. For an example, see Kooiman and Kloek (1985). The interference of this unboundedness phenomenon makes that one has to be very careful when interpreting estimation results. It is desirable to investigate the shape of the likelihood surface more closely in order to check the relevance of the estimates obtained. In this paper we intend to demonstrate that Bayesian methods can be used for that purpose.

Recently Van Dijk (1984) and Van Dijk, Kloek and Boender (this issue) have demonstrated the feasibility of numerical integration, hence Bayesian analysis, in relatively high-dimensional parameter space. They use Monte Carlo simulation methods for that purpose, based on importance sampling. We use two of their methods, named Simple Importance Sampling and Mixed Integration, to compute posterior first and second moments and marginal posterior density functions, both of the model parameters and of the average probability of the excess supply regime in the sample. These results, apart from being interesting by themselves, allow us to check whether the asymptotic maximum likelihood results adequately summarize the actual properties of the likelihood function, given the available set of data.

We present our switching model for retail services in section 2. In section 3 we derive the likelihood function, and in section 4 we discuss the methods of analysis employed. Estimation results and likelihood diagnostics by means of numerical integration methods are the subject of section 5. Section 6 concludes.

2. The model

The model we develop and discuss in this section expands on a model developed by Thurik and Koerts (1984a, b) and Thurik (1984) to explain floorspace productivity in retailing, i.e., the level of annual sales per square meter of available floorspace. The model assumes that the partitioning of floorspace into customer or selling area and remaining space plays a predominant role in the determination of the sales level. Differences in the partitioning of the floorspace reflect different marketing or operational strategies. A low share of selling area is found, for instance, in the traditional 'shop around the corner' with mainly counter service, where most goods are kept in stock and only few are displayed. A low share of remaining space is associated with modern self-service and cash-and-carry type shops, where most handling of goods takes place in the selling area and all goods are on display.

In the model both selling area and remaining space are treated as inputs in a production function for retail services. Input substitution accounts for changes in operational policy. The model assumes that shopkeepers partition their available floorspace in such a way that sales are maximized. Analytically the level of annual sales Q and the partitioning of available floorspace W in selling area C and remaining space R are determined by solving the following problem:

$$\max_{C,R} Q \quad \text{subject to} \quad Q \le Q^s(C, R; X), \tag{1}$$

$$R = W - C, \qquad 0 \le C \le W,$$

where X summarizes the exogenous variables of the model apart from W, and $Q^s(\cdot)$ is the possibility frontier or supply function for retail services. One may wonder whether it is appropriate to leave the demand side out, as Thurik (1984) does, and explain the level of sales from the supply side alone. Actually demand may not be large enough to sustain the solution derived from (1). In order to account for this possibility we introduce an additional demand constraint in the maximization problem (1):

$$Q \le Q^d(C; X), \tag{2}$$

where $Q^d(\cdot)$ is a demand function that we shall assume to be known by the shopkeeper. Thus we obtain a more general model that allows for 'demand-constrained' operations of retail establishments as well.

The general solution to the maximization problem (1) + (2) depends on the specification of the supply and demand functions. We employ the following beta-type supply function:

$$Q^{s}(C;W,X)=\beta(X)(C-\gamma)^{\pi\varepsilon}(W-C)^{(1-\pi)\varepsilon}, \tag{3}$$

for

$$0\leq\gamma\leq C,\quad 0\leq\pi\leq 1,\quad \varepsilon>0,\quad \beta(X)>0,$$

where we have suppressed the explicit dependence on R by substituting $R=W-C$. According to this equation supply is zero, i.e., no sales are possible, when remaining space is zero, or when selling area C does not surpass a threshold level γ independent of shop size, that we shall treat as a parameter to be estimated.[2] Other parameters to be estimated are the scale elasticity ε and the distribution parameter π. We postpone the specification of the shift factor $\beta(X)$ until section 5, where we present our estimation results, as it is immaterial for the general structure of our model on which we want to concentrate now. We refer to Thurik (1984) for a detailed justification of (3).

The demand function we employ is of the simple constant elasticity type:

$$Q^{d}(C;X)=\delta(X)(C-\gamma)^{\upsilon},\qquad \upsilon>0,\quad \delta(X)>0, \tag{4}$$

where we have included the same threshold γ as in eq. (3). The specification of $\delta(X)$ is again postponed until section 5. In view of the stochastic specification of our model, which is the subject of the next section, we have to impose the regularity condition $\upsilon\geq\pi\varepsilon$ on the demand elasticity υ. It guarantees[3] that multiple intersections of the supply and demand curves are excluded on the interval $\gamma<C\leq W$.

Given supply and demand equations (3) and (4) the solution to the sales maximization problem (1) + (2) takes one of two possible forms, depending on the relative position of the two curves. Fig. 1 depicts both possibilities. With demand large enough, as in fig. 1a, the optimum is obtained at the top of the $Q^{s}(\cdot)$ curve. When demand is too low to sustain this solution, the optimum is found at the intersection of the supply and demand curves, as in fig. 1b. In other words, in situations of excess supply shopkeepers tend to increase the share of selling area, and, thereby, decrease the service level (so-called 'trading

[2] Originally Thurik used a threshold for remaining space as well, which was consistently found to be zero, so we left it out.

[3] By comparing the derivatives of the right-hand sides of (3) and (4) with respect to C it can directly be checked that, given $\upsilon\geq\pi\varepsilon$, the demand curve intersects the supply curve from below for any intersection on the interval $\gamma<C\leq W$. For continuous functions this entails that at most one intersection can occur.

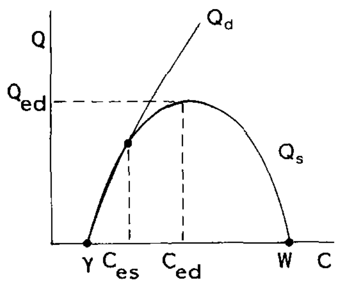

Fig. 1a. Excess demand;
$Q = Q_{ed}, C = C_{ed} \geq C_{es}$.

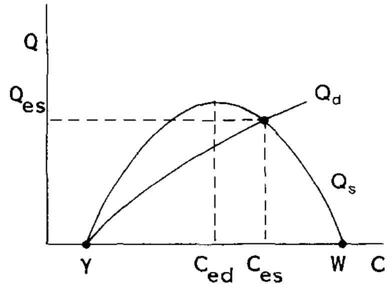

Fig. 1b. Excess supply;
$Q = Q_{es}, C = C_{es} > C_{ed}$.

down'). It can easily be seen from the figure that analytically the value of C follows as the maximum of C_{ed} and C_{es}, i.e., $C = \max(C_{ed}, C_{es})$, where C_{ed} is the solution to the first-order condition $\partial Q^s(\cdot)/\partial C = 0$, and C_{es} solves the equilibrium condition $Q^s(\cdot) = Q^d(\cdot)$. Moreover, it is immediately clear from the figure that the solution always lies on the supply curve. Consequently, our model for the endogenous variables Q and C reads as

$$\begin{aligned} &Q = Q^s(C; W, X), \\ &C = \max(C_{ed}, C_{es}), \\ &\partial Q^s(C_{ed}; W, X)/\partial C_{ed} = 0, \\ &Q^s(C_{es}; W, X) = Q^d(C_{es}; X), \end{aligned} \tag{5}$$

where the last two equations only serve to define the latent variables C_{ed} and C_{es} figuring in the maximum condition for C.

The type of model that we have obtained is known in the econometrics literature as a switching or minimum-condition model. Its canonical form obtains in the description of markets in disequilibrium, where transactions are assumed to equal the minimum of supply and demand, i.e., $Y = \min(Y^s, Y^d)$ with $Y^s = Y^s(P; X)$ and $Y^d = Y^d(P; X)$ as supply and demand functions, P being the price. The presence of an additional endogenous variable Q, that is not directly affected by the switch of regimes, makes our model analytically similar to a disequilibrium model with an additional price equation. The one-market disequilibrium model with endogenous prices has been described by Maddala and Nelson (1974), Goldfeld and Quandt (1975), Quandt (1982) and others. Rosen and Quandt (1978) give an empirical application to the U.S. labour market.

3. The likelihood function

Two types of stochastic specification have been discussed in the statistical literature on switching models based on the minimum condition. The first one goes back to Fair and Jaffee (1972) and Maddala and Nelson (1974). They assume that the level of transactions Y equals the exact minimum of unobserved stochastic supply and demand: $Y=\min(Y^{\mathrm{s}}, Y^{\mathrm{d}})$. The likelihood function for this model can be shown to be unbounded when variances tend to zero. To avoid this problem Ginsburgh et al. (1980) employ certainty equivalents and add an error term to the minimum condition: $Y=\min(\mathrm{E}Y^{\mathrm{s}}, \mathrm{E}Y^{\mathrm{d}})$ + error, where E is the mathematical expectations operator. The model for Y then effectively reduces to a non-linear regression model. As we find it hard to accept the implicit restriction in the latter approach that the supply and demand errors are identical, we prefer to employ the former one. For model (5) this implies that we add structural disturbance terms to the supply and demand equations (3) and (4). We use a multiplicative error specification to avoid heteroskedasticity and because it guarantees that Q^{s} and Q^{d} are zero for $C=\gamma$ and Q^{s} is zero for $C=W$ in the stochastic version of our model as well:

$$Q^{\mathrm{s}}=Q^{\mathrm{s}}(C; W, X)\exp(\varepsilon^{\mathrm{s}}), \tag{6}$$

$$Q^{\mathrm{d}}=Q^{\mathrm{d}}(C; X)\exp(\varepsilon^{\mathrm{d}}). \tag{7}$$

As is usually done we shall assume that the errors ε^{s} and ε^{d} are independently and identically normally distributed with zero means and variances σ_{s}^2 and σ_{d}^2.

It is not sufficient to include error terms in the supply and demand equations alone. Since we have two endogenous variables, Q and C, we need at least two error terms being 'active' for each regime. As under the excess demand regime both Q and C are determined from the supply function, we need an additional error term to be included in the determination of C_{ed}.[4] Using eq. (3) the first-order condition $\partial Q^{\mathrm{s}}(C_{\mathrm{ed}}; W, X)/\partial C_{\mathrm{ed}}=0$ yields $C_{\mathrm{ed}}=C_{\mathrm{ed}}(W)$, with $C_{\mathrm{ed}}(W)=\gamma+\pi(Q-\gamma)$. Aiming at a multiplicative error specification here as well we found it convenient to employ the following equation for C_{ed}:

$$(C_{\mathrm{ed}}-\gamma)/(W-\gamma)=\exp(-\phi), \tag{8}$$

[4]A similar procedure is common practice in production function studies, where an error term is introduced in the factor demand equations obtained from the first-order conditions for a maximum of (expected) profits, given the available technology.

where the error ϕ is restricted to the positive reals only, and $\mathrm{E}\exp(-\phi)=\pi$ is imposed in order to satisfy $C_{\mathrm{ed}}=C_{\mathrm{ed}}(W)$ on average. We shall assume that ϕ is independently[5] gamma distributed with density function $g(\phi;\alpha,\psi)=\Gamma(\alpha)^{-1}\psi^{-\alpha}\phi^{\alpha-1}\exp(-\phi/\psi)$, where $\Gamma(\alpha)$ is the (complete) gamma integral $\int_0^{\infty}z^{\alpha-1}\exp(-z)\,\mathrm{d}z$. As we want to impose zero probability density on the event $C=\gamma$, we shall impose $\alpha>1$. The scale parameter ψ has to be positive. It follows directly from the moment generating function of the gamma distribution, $M(t):=\mathrm{E}\exp(t\phi)=(1-t\psi)^{-\alpha}$, that the condition $\mathrm{E}\exp(-\phi)=\pi$ gives rise to the restriction

$$\alpha\log(1+\psi)=-\log\pi. \tag{9}$$

We shall use (9) to substitute for ψ so that the gamma distribution that we employ for ϕ only yields α as an additional parameter to be estimated.

The likelihood function of our model is derived from the joint density function of the endogenous variables Q and C. It is demonstrated in the *appendix* that the latter takes the following form:

$$h(Q,C)=\int_{\gamma}^{C}f(C,C_{\mathrm{es}},Q)\,\mathrm{d}C_{\mathrm{es}}+\int_{\gamma}^{C}f(C_{\mathrm{ed}},C,Q)\,\mathrm{d}C_{\mathrm{ed}}, \tag{10}$$

where $f(\cdot)$ is the joint density function of C_{ed}, C_{es} and Q as it can be derived from eqs. (6), (7) and (8) using the maximum condition $C=\max(C_{\mathrm{ed}},C_{\mathrm{es}})$. Instead of deriving the general form of $f(\cdot)$, it is more practical to work out the two right-hand-side terms separately and directly derive the required expressions from the 'special' version of the general model (5) that one obtains under each one of the two regimes.

We first need some notation. We have already introduced $g(\cdot;\alpha,\psi)$ to denote the gamma density function of ϕ. Let $G(\cdot;\alpha)$ denote the corresponding standardized ($\psi=1$) gamma distribution function. Similarly let $n(\cdot;\sigma)$ and $N(\cdot)$ denote the normal density function with zero mean and variance σ^2, and the standard normal distribution function respectively. We also define the residuals

$$e^{\mathrm{s}}:=q-q^{\mathrm{s}}(c;W,X),$$

$$e^{\mathrm{d}}:=q-q^{\mathrm{d}}(c;X),$$

$$p:=\log(W-\gamma)-\log(C-\gamma),$$

[5] One may doubt the independence of ϕ and ε^{s} since both errors account for discrepancies between our specification of the supply side and the data. Specification errors in supply equation (3) are likely to show up in both (6) and (8). To our defense we can point at common practice in production function studies that usually assume independence of 'technical' and 'allocative' inefficiency. The former refers to errors in the supply function *per se*, whereas the latter refers to errors in the demand equations derived from first-order conditions. Also, Thurik (1984, p. 43) reports correlation coefficients between the residuals of (6) and (8) that are less than 0.25 in absolute value.

where we have (partly) changed to logarithms, i.e., $q := \log Q$ and $c := \log C$. Notice that these residuals[6] are functions of the observables only.

Now in case of an *excess demand* we have $q = q^{\mathrm{s}}$, $c = c_{\mathrm{ed}}$, whereas c_{es} and q^{d} are unobserved. Taking logarithms in eqs. (6)–(8), and substituting for q^{s} and c_{ed} we directly obtain

$$\varepsilon^{\mathrm{s}} = q - q^{\mathrm{s}}(c; W, X) \quad (= e^{\mathrm{s}}), \tag{11a}$$

$$\varepsilon^{\mathrm{d}} = q^{\mathrm{d}} - q^{\mathrm{d}}(c; X), \tag{11b}$$

$$\phi = \log(W - \gamma) - \log(C - \gamma) \quad (= p). \tag{11c}$$

The regime only applies when $c_{\mathrm{es}} \leq c$, which, given the shape of our supply and demand curves, is equivalent to $q^{\mathrm{d}} \geq q$; compare fig. 1a. This, in turn, is equivalent to $\varepsilon^{\mathrm{d}} \geq e^{\mathrm{d}}$, as can be seen from the definition of e^{d} and eq. (11b). Consequently the first term in the joint density function of q and c is given by

$$h_{\mathrm{ed}}(q, c) = \int_{e^{\mathrm{d}}}^{\infty} f_{\mathrm{ed}}(q, c, \varepsilon^{\mathrm{d}}) \mathrm{d}\varepsilon^{\mathrm{d}},$$

where $f_{\mathrm{ed}}(\cdot)$ is the joint density function of q, c and ε^{d}. It can easily be obtained from (11) as follows. First factorize $f_{\mathrm{ed}}(q, c, \varepsilon^{\mathrm{d}})$ as $f_{\mathrm{ed}}^1(c) f_{\mathrm{ed}}^2(q|c) f_{\mathrm{ed}}^3(\varepsilon^{\mathrm{d}}|q, c)$. From (11c) we obtain $f_{\mathrm{ed}}^1(c)$ as $(C/(C-\gamma)) \times g(p; \alpha, \psi)$, the first factor being the Jacobian $\partial\phi/\partial c$ in absolute value. From (11a) $f_{\mathrm{ed}}^2(q|c)$ is directly obtained as $n(e^{\mathrm{s}}; \sigma_{\mathrm{s}})$, whereas $f_{\mathrm{ed}}^3(\varepsilon^{\mathrm{d}}|q, c)$ is simply $n(\varepsilon^{\mathrm{d}}; \sigma_{\mathrm{d}})$ as a consequence of the independence of ε^{d}, ε^{s} and ϕ. Integration over ε^{d} results in a factor $1 - N(e^{\mathrm{d}}/\sigma_{\mathrm{d}})$. Collecting factors, we end up with

$$h_{\mathrm{ed}}(q, c) = \frac{C}{C - \gamma} g(p; \alpha, \psi) n(e^{\mathrm{s}}; \sigma_{\mathrm{s}}) \left\{1 - N(e^{\mathrm{d}}/\sigma_{\mathrm{d}})\right\}.$$

Under *excess supply* we have $q = q^{\mathrm{s}}$, $q = q^{\mathrm{d}}$ from which c is implicitly determined as c_{es}, whereas c_{ed} is unobserved. As the alternative to (11) we now obtain

$$\varepsilon^{\mathrm{s}} = q - q^{\mathrm{s}}(c; W, X) \quad (= e^{\mathrm{s}}), \tag{12a}$$

$$\varepsilon^{\mathrm{d}} = q - q^{\mathrm{d}}(c; X) \quad (= e^{\mathrm{d}}), \tag{12b}$$

$$\phi = \log(W - \gamma) - \log(C_{\mathrm{ed}} - \gamma). \tag{12c}$$

The regime only applies when $c_{\mathrm{ed}} < c$, or, equivalently, $\phi > p$, so that the

[6] Only e^{s} is a proper residual since all observations, no matter the ruling regime, have to satisfy the supply function. The other two, e^{d} and p, can only be counted as observations of the error processes ε^{d} and ϕ to the extent that the correct regime applies.

second term of the joint density function of q and c is given by

$$h_{\mathrm{es}}(q,c)=\int_{p}^{\infty} f_{\mathrm{es}}(q,c,\phi)\,\mathrm{d}\phi,$$

where $f_{\mathrm{es}}(q,c,\phi)$ is the joint density function of q, c and ϕ. Again we factorize $f_{\mathrm{es}}(q,c,\phi)$ as $f_{\mathrm{es}}^{1}(q,c)f_{\mathrm{es}}^{2}(\phi|q,c)$. We obtain $f_{\mathrm{es}}^{1}(q,c)$ by changing variables from the joint normal density function of ε^{s} and ε^{d} using (12a) and (12b). This yields a factor $n(e^{\mathrm{s}};\sigma_{\mathrm{s}})n(e^{\mathrm{d}};\sigma_{\mathrm{d}})$ and a Jacobian factor $|\partial q^{\mathrm{d}}(c;X)/\partial c-\partial q^{\mathrm{s}}(c;W,X)/\partial c|$, which, using eqs. (3) and (4), equals $(v-\pi\varepsilon)C/(C-\gamma)+(1-\pi)\varepsilon C/(W-C)$. The density function $f_{\mathrm{es}}^{2}(\phi|q,c)$ is $g(p;\alpha,\psi)$, so that the integration over ϕ results in a factor $1-G(p/\psi;\alpha)$. Collecting factors, we find

$$h_{\mathrm{es}}(q,c)=\left\{\frac{(v-\pi\varepsilon)C}{C-\gamma}+\frac{(1-\pi)\varepsilon C}{W-C}\right\}n(e^{\mathrm{s}};\sigma_{\mathrm{s}})n(e^{\mathrm{d}};\sigma_{\mathrm{d}})$$
$$\times\{1-G(p/\psi;\alpha)\}.$$

Adding a subscript i to indicate observations, the likelihood function can now be formulated as

$$L=\prod_{i=1}^{N}\{h_{\mathrm{ed}}(q_i,c_i)+h_{\mathrm{es}}(q_i,c_i)\},\tag{13}$$

where N is the total number of observations.

Although the likelihood function obtained looks rather awkward its general structure is transparent: for each observation we obtain a weighted sum of the joint density functions of q and c under each of the two possible regimes. The weights are the unconditional probabilities that the condition under which the regime applies is satisfied. These probabilities do not add to unity, contrary to the conditional probabilities that a *given* observation on q and c has been generated under the excess supply regime or the excess demand regime.[7] It is shown in the appendix that these probabilities are equal to

$$\begin{aligned}P_{\mathrm{es}}(q,c)&=h_{\mathrm{es}}(q,c)/(h_{\mathrm{es}}(q,c)+h_{\mathrm{ed}}(q,c)),\\P_{\mathrm{ed}}(q,c)&=h_{\mathrm{ed}}(q,c)/(h_{\mathrm{es}}(q,c)+h_{\mathrm{ed}}(q,c)).\end{aligned}\tag{14}$$

[7] We refer to Gersovitz (1980) and Kiefer (1980) for a discussion of the issue of regime classification in disequilibrium models.

Once the model has been estimated we can use these expressions to obtain estimates of the two regimes probabilities for each observation in the sample. This is most interesting from a policy point of view as it provides us with an estimate of the extent to which the sector under study operates supply-constrained or demand-constrained. It also allows us to check whether a short-cut is possible based on the assumption that all observations effectively come from one regime only.

We conclude this section with a few remarks on the unboundedness of our likelihood function (13). In the appendix we show that the likelihood becomes unbounded in switching models of this type when variances tend to zero and other parameters take such values that completely one-sided samples are implied. For the present model this occurs when σ_d tends to zero and the demand-side parameters take such values that all observations satisfy $Q \leq Q^d(C; X)$, with equality for one observation at least. The likelihood also becomes unbounded in the opposite situation where the variance of ϕ tends to zero and the supply-side parameters take such values that all observations satisfy $C \geq C_{ed}(W)$, with equality for one observation at least. The condition $C \geq C_{ed}(W)$ implies that, for arbitrary but given γ, π has to be fixed at the smallest value of $(C-\gamma)/(W-\gamma)$ in the sample. Moreover, the variance of ϕ is $\alpha\psi^2$. As we have imposed $\alpha > 1$ it can only be equal to zero when ψ equals zero. From (9) it is clear then that in this situation the likelihood can only become unbounded when α tends to infinity.

4. Estimation methods

Estimation of our model proceeds in two steps. First we compute classical maximum likelihood estimates of the model parameters. An estimate of the covariance matrix is obtained in the usual way as minus the inverse of the Hessian matrix of the log-likelihood function evaluated at the point estimates obtained. Secondly, we check whether these asymptotic results adequately summarize the properties of our actual likelihood function. For this purpose we use a Bayesian methodology applying numerical integration by means of Monte Carlo methods. We obtain posterior means and the posterior covariance matrix of the parameters of the model. We also compute marginal posterior density functions for individual parameters of interest. The methods employed allow for the computation of marginal posterior density functions of arbitrary functions of the parameters as well. We exploit this feature to compute the posterior mean and variance, and the posterior density function of the average probability of the excess supply regime for the observations in our sample. Thus we get an idea about the information content of the data with respect to the regime distribution.

4.1. Maximum likelihood

The first step in our analysis consists of finding the maximum of the likelihood function (13). The main problem is, of course, that the likelihood function is known to become unbounded when either σ_d goes to zero or α tends to infinity, and other parameters take suitable values as well. Compare the discussion at the end of the preceding section. Actually we are not really interested in the properties of the likelihood in this region of the parameter space as zero variances of the errors have a priori probability (density) zero in the rather simple and highly stylized model we analyse.[8] What we are really looking for is a (local) maximum located in a domain in parameter space that excludes this anomalous region. Hartley and Mallela (1977) have shown that the usual asymptotic properties of maximum likelihood apply to such a point, provided the true parameter vector is located in the interior of the domain considered.

This leaves us with the need to include a penalty function such that the unboundedness region is effectively suppressed without affecting the shape of the likelihood in a more realistic domain of the parameter space. One possible way to proceed is to impose a strictly positive lower bound on σ_d and an upper bound on α, but it is rather difficult to make up one's mind as to what values are acceptable. So we adopted another strategy. When the likelihood becomes unbounded we obtain an almost entirely one-sided sample with all observations assigned to one regime with probability 1, except for one or two observations that are assigned with probability 1 to the other regime. From eq. (14) it is clear that regime probabilities are obtained without extra costs as a by-product when evaluating the likelihood function. This suggests to constrain the range of acceptable values for the average probability of, say, the excess supply regime in the sample. So, in order to keep the optimization path away from the unboundedness region in the parameter space, we added the following penalty function to the log-likelihood function:

$$
\begin{aligned}
P &= b(1-\bar{P}_{es}/B_l)^2/\bar{P}_{es} && \text{iff} \quad \bar{P}_{es} \le B_l, \\
&= b\frac{\left(1-\dfrac{1-\bar{P}_{es}}{1-B_u}\right)^2}{1-\bar{P}_{es}} && \text{iff} \quad \bar{P}_{es} \ge B_u, \\
&= 0 && \text{elsewhere,}
\end{aligned}
\tag{15}
$$

where b is a scaling factor that serves to vary the intensity of the penalty, and B_l and B_u are lower and upper bounds for the average probability of the

[8] Note that, quite contrary to this statement, the usual 'non-informative' prior $p(\sigma) = \sigma^{-1}$ does give high credit to this state of nature.

excess supply regime in the sample

$$\bar{P}_{\mathrm{es}} = N^{-1} \sum_{i=1}^{N} P_{\mathrm{es}}(q_i, c_i); \tag{16}$$

compare eq. (14). Note that this penalty function is continuously differentiable with respect to $\bar{P}_{\mathrm{es}}$, which in turn is a complicated, but continuously differentiable function of the model parameters. This guarantees that the function to be maximized has a continuous gradient, which is desirable in view of the optimization routine that we used.[9]

4.2. Likelihood diagnostics

From now on we shall reinterpret our likelihood function as the kernel of a joint posterior density function of the model parameters. Obviously, this entails the use of a constant prior for our further analysis. The switch to a Bayesian methodology allows us to investigate global properties of the likelihood surface by means of numerical integration methods. Van Dijk (1984) has demonstrated that numerical integration, hence Bayesian analysis, is feasible even in relatively high-dimensional parameter space, when using Monte Carlo simulation methods. Posterior means and (co-)variances and marginal posterior density functions can be computed, both of parameters of interest and of interesting functions of those parameters, like multipliers, adjustment speed, etc.

In this subsection we shall briefly describe the methodology involved. More details can be found in Van Dijk (1984) and in Van Dijk, Kloek and Boender (in this volume).

Let $p(\theta)$ be the kernel of a joint posterior density function of a parameter vector θ. Suppose we want to investigate the distribution of some scalar function $f(\theta)$ of θ. Evaluation of, say, its first two moments requires the computation of

$$\mathrm{E}f(\theta) = \frac{\int f(\theta) p(\theta)\,\mathrm{d}\theta}{\int p(\theta)\,\mathrm{d}\theta}, \tag{17}$$

$$\mathrm{E}f(\theta)^2 = \frac{\int f(\theta)^2 p(\theta)\,\mathrm{d}\theta}{\int p(\theta)\,\mathrm{d}\theta}, \tag{18}$$

[9] We used NAG-liberty routine E04JBF, with numerical first derivatives. One optimization run takes between 500 and 1000 functions evaluations, which costs between 50 and 100 seconds cpu on a CDC Cyber 170-855. To evaluate $N(z)$ and $\Gamma(\alpha)$ we used NAG routines S15ABF and S14ABF, respectively. Values for the incomplete gamma integral $G(z;\alpha)$ were obtained by means of an expansion presented by Lau (1980) (algorithms AS 147).

where the integration is over the entire domain of θ. The posterior density function of $f(\theta)$ can be approximated once we are able to evaluate posterior probabilities of the type $\Pr(a_1 < f(\theta) \leq a_2)$, where (a_1, a_2) is a small interval in the range of possible values for $f(\theta)$. These probabilities, however, can easily be seen to be equal to the expectation of a dummy function $D(\theta)$ defined as

$$\begin{aligned} D(\theta) &= 1 \quad \text{iff} \quad a_1 < f(\theta) \leq a_2, \\ &= 0 \quad \text{elsewhere.} \end{aligned} \tag{19}$$

Consequently, the general numerical problem involved is how we can compute integrals of the type

$$\mathscr{T}_g = \int g(\theta) p(\theta) \, \mathrm{d}\theta,$$

for arbitrary functions $g(\theta)$.

The integration method we employ is based on so-called *importance sampling*. Let $I(\theta)$ be a density function associated with a probability distribution from which we can *easily* generate (pseudo-)random drawings by means of a computer. Then

$$\mathscr{T}_g = \int g(\theta) \frac{p(\theta)}{I(\theta)} I(\theta) \, \mathrm{d}\theta = \mathrm{E}_I \left\{ g(\theta) \frac{p(\theta)}{I(\theta)} \right\},$$

where E_I denotes the mathematical expectation with respect to $I(\theta)$. From the law of large numbers it follows directly that $\mathscr{T}_g$ can consistently be estimated from a random sample θ_j $(j = 1, \ldots, J)$ drawn from a distribution with density $I(\theta)$ as

$$\hat{\mathscr{T}}_g = J^{-1} \sum_{j=1}^{J} g(\theta_j) \frac{p(\theta_j)}{I(\theta_j)}. \tag{20}$$

We call $I(\theta)$ the importance function as it specifies the density of the sampling process for each point in the domain of integration, i.e., its relative importance. The numerical precision of the estimate (20) crucially depends upon the variance of $g(\theta)p(\theta)/I(\theta)$. Therefore it is desirable to select an importance function that approximates the posterior density function as closely as possible. Several alternative methods of Monte Carlo integration originate in different principles to select or construct an appropriate importance function. We shall discuss two methods that have proved to be successful in some applications.

The first method is Simple Importance Sampling (SIS) where we use a member of the multivariate Student-t family of density functions as our importance function. The multivariate normal density function is a limiting member of this family. Maximum likelihood estimates are used to specify the values of the location and scale parameters. Tail behavior can be varied by choosing different values for the degree of freedom parameter. As we expected the posterior density function to have relatively fat tails we opted for one degree of freedom in the present application of the method. The actual integration results suggest that a somewhat larger value might have been more appropriate, though.

The other technique we employ is Mixed Integration (MIN). Its distinctive feature is that it employs a 'mixture' of classical numerical quadrature and importance sampling. Importance sampling is used to generate directions in parameter space, whereas for each direction a one-dimensional classical integration step is performed. Contrary to SIS, which is based on a symmetric importance function, MIN is robust towards asymmetric tail behaviour, or multivariate skewness, of the posterior density function.

Actually, the MIN technique proceeds by changing variables in the integrand of $\mathscr{T}_g$, using a transformation of the k-vector θ of parameters into a pair (ρ, η). The $(k-1)$-vector η represents the direction of the vector $\theta - \theta_0$, θ_0 being the posterior mode. The scalar ρ satisfies $\rho^2 = (\theta - \theta_0)'V^{-1}(\theta - \theta_0)$, V being the importance covariance matrix. A sign convention for ρ is added to ensure that the transformation is one-to-one. The actual transformation employed involves a Jacobian determinant $|\rho|^{k-1}|J(\eta)|$, where, as indicated, the factor $|J(\eta)|$ only depends on η. Letting $\theta(\rho, \eta)$ denote the inverse transformation, we obtain for $\mathscr{T}_g$

$$
\begin{aligned}
\mathscr{T}_g &= \int g(\theta(\rho, \eta))\, p(\theta(\rho, \eta))\, |\rho|^{k-1} |J(\eta)|\, \mathrm{d}\rho\, \mathrm{d}\eta \\
&= \int_\eta \left\{ \int_\rho g(\theta(\rho, \eta))\, p(\theta(\rho, \eta))\, |\rho|^{k-1}\, \mathrm{d}\rho \right\} |J(\eta)|\, \mathrm{d}\eta \\
&= \mathrm{E}_{|J|} \left\{ \int_\rho g(\theta(\rho, \eta))\, p(\theta(\rho, \eta))\, |\rho|^{k-1}\, \mathrm{d}\rho \right\},
\end{aligned}
$$

where $\mathrm{E}_{|J|}$ denotes the mathematical expectation with respect to $|J(\eta)|$ considered as the kernel of a density function for η. From the transformation employed it can easily be shown that random drawings η_j can be generated from a distribution which has a density function proportional to $|J(\eta)|$ by simply generating random drawings θ_j from a multivariate normal with mean vector θ_0 and covariance matrix V, and then applying the transformation involved. Accordingly, we can estimate $\mathscr{T}_g$ using J random drawings η_j as[10]

$$\hat{\mathscr{T}}_g \propto J^{-1} \sum_{j=1}^{J} \left\{ \int_\rho g(\theta(\rho,\eta_j)) p(\theta(\rho,\eta_j)) |\rho|^{k-1} \mathrm{d}\rho \right\}.$$

For each random drawing η_j we have to compute the value of the integral over ρ, for which we use a 16-point Gaussian quadrature. Numerical convergence is checked by running separate integrations over subranges of the domain of ρ. Thus the actual number of function evaluations per integration step is considerably larger than 16. The numerical quadrature step complicates the computation of marginal posterior density functions according to eq. (19). The value obtained from the integral over ρ has to be 'redistributed' over the various intervals (a_1, a_2) that partition the domain of each of the parameters. A crude but effective solution consists of assigning the contribution of each of the parameter points at which the integrand is evaluated to the interval in which this parameter point happens to fall.

5. Results

We have estimated our model with data from a survey of Dutch independent supermarkets and superettes conducted in 1979 by the Research Institute for Small- and Medium-Sized Business EIM.[11] The sample consists of 215 shops with floorspace ranging from approximately 110 to 1600 m^2. For these establishments a large number of operational, financial and environmental variables have been observed. For this study we only make use of a limited number of these variables. Our selection was primarily based on the results earlier obtained by Thurik (1984), where the reader can also find a detailed account of the available variables.

We first discuss the specification of the shift factors $\beta(X)$ and $\delta(X)$ of our supply and demand functions (3) and (4). Some preliminary exercises indicated that the following specification performed rather well:

$$\beta(X) = \exp(\beta_0)(1+M)H^{\beta_1}, \tag{21}$$

$$\delta(X) = \exp(\delta_0 + \delta_1 F)(1+M)^{1+\delta_2}, \tag{22}$$

where M is the fractional gross margin $(Q - PV)/PV$, PV being the purchasing value of sales Q, H is occupancy costs per square meter, and F is the relative share of sales of fresh products, as e.g., dairy products, bread, fruits, and vegetables. Meat and meat products are not included in this variable.

[10] Proportionality is sufficient as we always compute ratios of integrals, so that the integration constants cancel. Compare eqs. (17) and (18).

[11] The data can be obtained from the authors upon request.

The factor $(1+M)$ in (21) is a proxy for prices, that we do not observe. Being equal to Q/PV its role is to transform the value of sales Q into its volume, that we assume proportional to the purchasing value PV. In (22) the factor $1+M$ is present for the same reason, but also because we expect prices to influence the level of demand. Consequently, we interpret the parameter δ_2 as a price elasticity, expected to be negative. It is customary in retail productivity studies to assume that productivity increases with factor prices.[12] High factor costs urge the shopkeeper to exploit his resources efficiently. Housing being a production factor in the retail industry, we have included occupancy costs in (21), where it serves as a proxy for efficiency. We expect the parameter β_1 to be positive. Occupancy costs are likely to be correlated with the quality of the site too. Thus one can argue equally well that H has to be included in (22), where it serves as a proxy for attractiveness, i.e., environmental factors influencing demand. Including H in (22) instead of (21) we obtained entirely unacceptable estimates, though. We have also tried to include H in both (21) and (22). Constraining the elasticities to be non-negative, we invariably obtained zero elasticity of demand, so we continued with the specifications according to (21) and (22). We expect the parameter δ_1 in (22) to be positive as the availability of fresh products is likely to exert a positive influence on demand.

Confronting our model with the available data the first step to be taken was to find the maximum of the likelihood function – or equivalently: the posterior mode – and to evaluate the inverse Hessian matrix of the log-likelihood at the optimum. Initially optimization runs did not properly converge due to misspecification of $\beta(X)$ and $\delta(X)$, and the presence of outliers. During the process of selecting eqs. (21) and (22) we also deleted 7 out of the 215 available observations. These showed up as clear outliers, either because of an extremely bad fit on the supply curve (3), or because of a very low likelihood value (13). The first two columns of table 1 give the optimisation results obtained with the remaining 208 sample points. Taking $B_1 = 1 - B_u = 0.1$ and $b = 50$ in penalty function (15) a unique local maximum obtains. According to the implied estimates of the regime probabilities the average probability of the excess supply regime in the sample is 0.242. Although this implies that the penalty is not active at the final point, it does play an important role during the iterations leading to this point: leaving out the penalty function we repeatedly (but not invariably) end up in the unboundedness region with the optimization routine failing to converge.

Now turning to the point estimates obtained, we first notice that we have deleted the threshold γ figuring in eqs. (3) and (4) since it invariably ended up at the imposed lower bound of zero. Thurik (1984) reports the same result for

[12] For references see, for instance, *Journal of Retailing*, Special issue on Productivity, Vol. 60, no. 3, Fall 1984.

Table 1

Parameter estimates (standard errors in parentheses).

Par.	Eq.	ML	ML ($\delta_2=0$)	MIN	SIS	Thurik
β_0	(21)	2.942 (0.261)	2.943 (0.261)	2.945 (0.264)	2.950 (0.265)	3.046
β_1	(21)	0.749 (0.052)	0.749 (0.052)	0.748 (0.053)	0.747 (0.053)	0.724
ε	(3)	0.865 (0.026)	0.865 (0.026)	0.865 (0.026)	0.866 (0.027)	0.857
π	(3)	0.642 (0.015)	0.643 (0.015)	0.644 (0.017)	0.644 (0.016)	0.683
σ_s	(6)	0.223 (0.011)	0.223 (0.011)	0.226 (0.011)	0.226 (0.011)	0.220
δ_0	(22)	6.524 (0.296)	6.451 (0.158)	6.487 (0.182)	6.493 (0.183)	—
δ_1	(22)	1.390 (0.540)	1.377 (0.531)	1.363 (0.617)	1.351 (0.598)	—
δ_2	(22)	−0.362 (1.24)	0.000	0.000	0.000	—
υ	(4)	0.910 (0.050)	0.907 (0.049)	0.906 (0.056)	0.904 (0.059)	—
σ_d	(7)	0.195 (0.027)	0.195 (0.027)	0.209 (0.034)	0.210 (0.035)	—
α	(9)	5.492 (0.633)	5.503 (0.633)	5.558 (0.647)	5.568 (0.633)	5.207
$\log L$		−317.913	−317.955			−332.987
$\bar{P}_{es}$		0.242	0.240	0.227 (0.066)	0.225 (0.063)	—

this particular data set. The other supply-side parameters take plausible values and appear to be fairly well determined. There is a strong effect of occupancy costs on sales performance ($\beta_1=0.75$), on which we have already commented. There are slight but significant diseconomies of scale ($\varepsilon<1$) and on average sales performance is maximal when about two-thirds of the available area is selling area ($\pi=0.66$). The demand side appears to be somewhat less well determined, except for the elasticity with respect to selling area, υ, which is close to unity. The price elasticity δ_2 has a very large standard error, probably because gross margin is a bad proxy for selling prices. We have deleted it in our further experiments in order to economize on the size of the model. It can be checked that the effect is negligible by comparing the first two columns of the table. The share of fresh products sold, with elasticity $\delta_1 F$, shows up relatively weak as well. The null hypothesis of no effect is rejected at the 95%

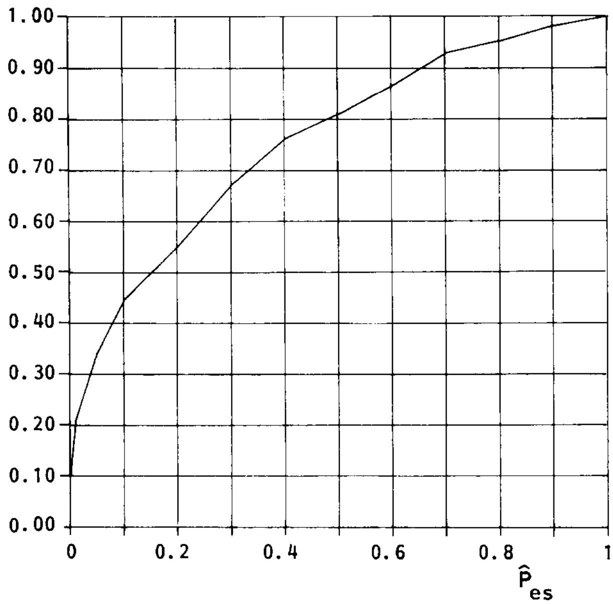

Fig. 2. Cumulative distribution of estimates of P_{es} in the sample.

level on a likelihood ratio test, though, twice the logarithm of the likelihood ratio being equal to 5.34.

We have tried to improve upon the specification of our demand equation by including a locational dummy in (22) indicating whether the establishment is located at a large shopping centre or not. This proved to be insignificant, twice the logarithm of the likelihood ratio being only 0.90. We have also included a dummy indicating whether or not the establishment contains a butcher's shop, but this consistently ended up at the imposed lower-bound of zero.

Finally, considering the error terms of our model, the standard deviations of the supply and demand errors ε^s and ε^d are about 20%, which is acceptable for cross-sectional data. The same figure obtains for the error ϕ in eq. (8) for the optimal partitioning of the floorspace. Its standard deviation can easily be obtained from eq. (9) and the estimates for α and π as $100\alpha^{\frac{1}{2}}\psi = 19.7\%$.

Now turning to the regime probabilities, fig. 2 displays the cumulative distribution of our estimates of P_{es} in the sample. We have already referred to the average value of these estimates, which is 0.24. It can be checked from the figure that the median value is about 0.15, and that more than 80% of the observations have a probability of excess supply less than one half. This illustrates that according to these estimates the majority of the observations are strongly supply determined. This may partly explain why the demand side is not so well determined. It also entails that Thurik's (1984) model, which assumes excess demand for all observations, may be an acceptable approximation. This is reflected in the last column of table 1, where we have reestimated our supply side under this assumption. Point estimates are only slightly different. The demand side is not identified when all observations are assumed to be supply determined.[13] Neither can log-likelihood values be compared in the usual way since the models are not properly nested.

It is possible, in principle, to compute estimates of the standard error of $\bar{P}_{\mathrm{es}}$ by evaluating the square root of $(\partial\bar{P}_{\mathrm{es}}/\partial\theta)'V_\theta(\partial\bar{P}_{\mathrm{es}}/\partial\theta)$, where V_θ is the estimated covariance matrix of the parameters. As the programming of the gradient vector of $\bar{P}_{\mathrm{es}}$ is not a trivial task, we prefer to deal with matters of precision in the context of the Bayesian approach, to which we shall turn now.

We gain further insight in the shape of the likelihood surface once we shift from a classical maximum likelihood (ML) approach to a Bayesian methodology. For that purpose we first have to specify the prior information that we want to use. As we are interested in the shape of the likelihood function *per se* we opt for uniform priors on relatively wide, but bounded intervals for all our parameters. The bounds that we employ determine the domain in parameter space where we perform integrations using SIS and MIN techniques. They have been selected using the earlier ML results and some preliminary integration runs. The domain of integration was further restricted by imposing $\delta_1 \geq 0$, and the regularity condition $\upsilon > \pi\varepsilon$ that we discussed in section 2. The domain includes values for the demand-side parameters for which we have an excess demand for all observations,[14] i.e., $Q \leq Q^{\mathrm{d}}(C;\, X)$, with equality for at least one observation. In this situation the likelihood increases without bound when σ_{d} goes to zero, which is the lower bound we employ for this parameter.

Obviously it is very dangerous to perform numerical integration when the integrand has a pole in the domain of integration. We even cannot be sure that the expectations we try to estimate exist. One possible way to proceed is to get rid of the pole by restricting σ_{d} to be strictly positive. Then, of course, the sensitivity of the final results with respect to the actual choice of the lower bound for σ_{d} becomes the central issue. It can be addressed by repeating the integrations for different choices. This, however, is very costly, so we choose to take the more risky approach and check during our integration runs whether any signs of actual problems of this type could be detected. Both with SIS and MIN we have checked for outliers, i.e., random drawings that contribute more

[13] We can assume that all observations satisfy the demand equation as well, but then we obtain an equilibrium model, and this entails the necessity to introduce a third endogenous variable in the model, i.e., one that adjusts fast enough to close temporary disequilibria.

[14] Since we use a bounded interval for α our domain of integration excludes unboundedness of the likelihood function originating in the opposite event of a general excess supply. Compare the discussion at the end of section 3. We have not checked the sensitivity of the integration results with respect to the value of the upperbound for α, as the results obtained did not indicate any influence whatsoever.

than expected to the value of the integral obtained. For the same purpose we have also monitored the convergence of the integration runs, both with SIS and MIN. Finally, we have tested for the occurrence of multiple modes when integrating along the directions randomly selected in the context of the MIN procedure. On these checks we have not obtained even the slightest indication that unboundedness problems might interfere. Unimodality was always confirmed, integration runs always converged fast and smoothly, and the distribution of the values of the integrand obtained for all random drawings was always quite acceptable. Consequently we are inclined to conclude that either the mass involved in the unboundedness region is negligible, or, when this is not the case, this region is sufficiently isolated and concentrated at the boundary of the domain of integration to cause no practical problems.

Now turning to the integration results, the third and fourth columns of table 1 present the posterior means and standard deviations obtained with the MIN and SIS methods, respectively. The methods estimate the same expectations, so differences in the results are only due to sampling variance in the Monte Carlo procedure. This variance is easily computed for SIS, whereas it can be approximated for MIN; compare Van Dijk (1984, subsection 3.8). The differences between the MIN and SIS estimates of the posterior means fall within 2σ bounds so computed. The MIN results are based upon 2000 random directions, with on average 90 function evaluations per quadrature for each of the directions obtained. SIS results are based on 50,000 function evaluations. As the methods converge relatively fast acceptable results are already available after 5–10,000 function evaluations.

Comparing the integration results with the ML estimates the similarity is striking. For the supply-side parameters the results are virtually identical. The asymptotic standard errors of the ML estimates of the demand-side parameters are consistently somewhat too low as compared to the (exact!) posterior standard deviations. Point estimates are close together here as well. Mode and mean being so close together, the posterior density function and hence the likelihood function, is likely to be almost symmetrical. This is confirmed by the shape of the marginal posterior density functions of the model parameters. Fig. 3 gives some examples, that are representative for the general pattern we found. The apparent symmetry of the likelihood function explains why SIS, which is based on a symmetrical importance function, performs so well. According to the estimated sampling variance it is even slightly more efficient than the MIN technique on the present model. In conjunction with the fast convergence of both methods, i.e., small sampling errors, this strongly indicates that the likelihood surface has a regular shape that can adequately be represented by a member of the multivariate Student-t family of density functions.

We have also computed integration results for the average probability of the excess supply regime, $\bar{P}_{\mathrm{es}}$. Its marginal density function is depicted in fig. 4. According to this result the data strongly reject the excess supply hypothesis

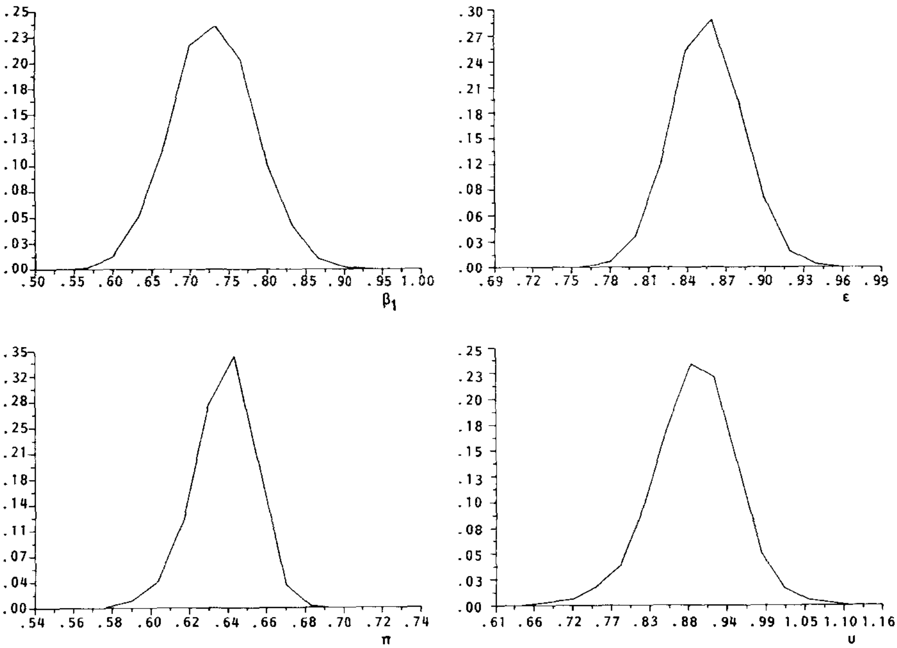

Fig. 3. Marginal posterior density functions for β_1, ε and π [eqs. (3) and (21)] and for v [eq. (4)].

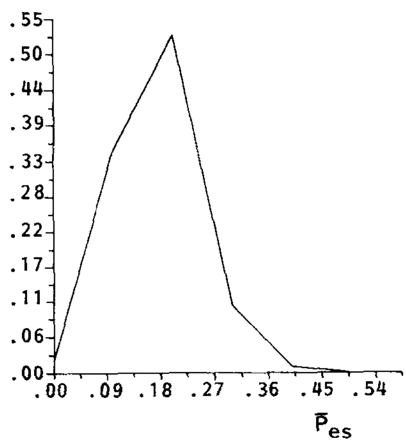

Fig. 4. Marginal density function of $\overline{P}_{es}$.

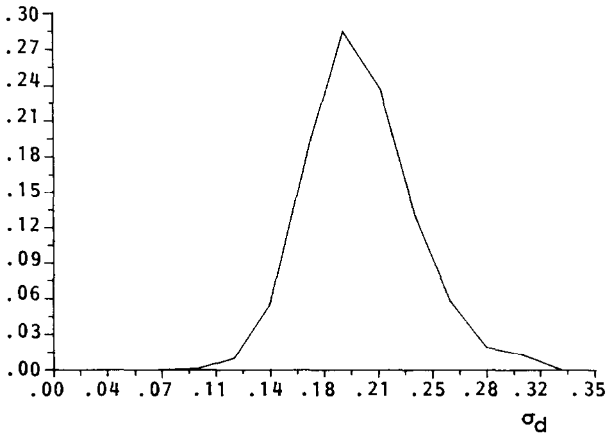

Fig. 5. Marginal posterior density function of σ_{d}.

for the majority of the observations, the probability density being practically zero for values larger than one half. The skewness of the density function is reflected in the fact that the mean is slightly smaller than the mode. A nice feature of the integration methodology employed is that we can get estimates of the precision practically without additional costs. The standard deviation of $\bar{P}_{\mathrm{es}}$ is estimated as 0.065. This confirms the idea that, given the specification of our supply and demand sides, the data are quite informative with respect to the regime distribution.

We conclude our discussion of the estimation results with fig. 5 which gives the marginal posterior density function of σ_{d}. It shows that values for σ_{d} less than 0.05 have a probability density practically equal to zero. This is in accord with the finding that unboundedness problems seem to play no role whatsoever in the integration results that we have presented.

6. Conclusion

From the statistical point of view our switching model of retailing services performs reasonably well. We have, of course, made some preselection, both in terms of model specification and data, but not more so than it is customary in empirical econometric work. Contrary to what we had expected, the likelihood surface seems to have a very regular and symmetrical shape, so that the standard asymptotic ML results are fully adequate as a summary of the sample information, given our model specification.

Unboundedness of the likelihood function, that can be shown to exist in our model, did not interfere, neither in the classical 'optimization', nor in the

Bayesian 'integration' stages of the estimation process. Apparently the 'spikes' associated with this phenomenon are sufficiently concentrated and isolated at the boundary of parameter space not to frustrate the numerical procedures that we have used for the diagnosis of the likelihood function of our model.

The integration methods that we have used, i..e, Simple Importance Sampling and Mixed Integration, perform very well. This is not a surprise in view of the regularity of the likelihood surface. It is still an open question whether the techniques perform equally well in a less friendly environment.

Now turning to the model that we have investigated, its main purpose was to help create a framework for further research into the influence of environmental factors on floorspace productivity. More knowledge is needed in this field in view of the desire of EIM to build and maintain a decision support system for retailers and consultants. The presence of both a supply side and a demand side in the model also allows us to estimate the degree of over-capacity for branches in the retail industry. This may be valuable from a policy point of view.

The estimation results allow for two main conclusions. The first is that occupancy costs are a supply factor, not a demand factor. It is a supply factor, probably, because the efficiency of the shopkeeper is positively correlated with the factor prices he has to pay: only efficient producers can afford to employ expensive resources. It is not a demand factor, probably, because occupancy costs are not a good proxy for environmental factors influencing demand. The second conclusion is that there is no drastic overcapacity in the small Dutch independent grocery trade in 1979. Given the observations in our sample the average probability of excess supply is estimated as 24% with a standard deviation of $6\frac{1}{2}$%. In view of the fact that small independent grocers have the least competitive power in the grocery trade, it is likely that in 1979 no overcapacity occurred in the Dutch grocery trade as a whole.

Appendix

Some properties of the minimum of two random variables

This appendix is included for expository purposes mainly. To accommodate the non-technical reader we use only elementary tools of mathematical statistics. The results derived are not new, similar derivations can be found in Maddala and Nelson (1974), Goldfeld and Quandt (1975), Quandt (1982), Maddala (1983) and others.

We shall deal with the canonical form of the model for a market in disequilibrium, where transactions y are the minimum of stochastic supply y^{s} and demand y^{d} with joint density function $g(y^{\mathrm{s}}, y^{\mathrm{d}})$. We shall first derive the density function of y. Then we obtain expressions for the conditional regime probabilities $\Pr(y^{\mathrm{d}} \leq y^{\mathrm{s}} | y)$ and $\Pr(y^{\mathrm{s}} < y^{\mathrm{d}} | y)$. We shall also demonstrate the unboundedness of the likelihood function associated with a random sample of observations on y. Finally, we show that the introduction of an additional endogenous variable that is not directly affected by the regime switch is straightforward.

We directly obtain the *density function* $f(y)$ by differentiation of the distribution function $F(Y) := \Pr(y \leq Y)$. As the two regimes are disjoint we have

$$F(Y) = \Pr\left(y^{d} \leq Y \wedge y^{d} \leq y^{s}\right) + \Pr\left(y^{s} \leq Y \wedge y^{s} < y^{d}\right)$$
$$= \int_{-\infty}^{Y} \int_{y^{d}}^{\infty} g(y^{s}, y^{d})\, \mathrm{d}y^{s} \mathrm{d}y^{d} + \int_{-\infty}^{Y} \int_{y^{s}}^{\infty} g(y^{s}, y^{d})\, \mathrm{d}y^{d} \mathrm{d}y^{s}.$$

Differentiation with respect to Y, and subsequent replacement of Y by y yields the density function $f(y)$,

$$f(y) = \left. \frac{\partial F(Y)}{\partial Y} \right|_{Y=y} = f^{\mathrm{es}}(y) + f^{\mathrm{ed}}(y), \tag{A.1}$$

where

$$f^{\mathrm{es}}(y) := \int_{y}^{\infty} g(y^{s}, y)\, \mathrm{d}y^{s}, \tag{A.2}$$

$$f^{\mathrm{ed}}(y) := \int_{y}^{\infty} g(y, y^{d})\, \mathrm{d}y^{d}. \tag{A.3}$$

Notice that in each of the two terms the domain of integration consists of the range of admissible values of the unobserved side of the market, given the observed level of transactions y. In case of a 'maximum' condition, as in the main text, this entails that integration bounds will be $-\infty$ and y instead of y and ∞.

Now turning to the *regime probabilities* we shall derive the following property:

$$\Pr\left(y^{d} \leq y^{s} | y\right) = f^{\mathrm{es}}(y)/f(y). \tag{A.4}$$

For that purpose we introduce the auxiliary random variable $z := y^{d} - y^{s}$. Let $h(y, z)$ be the joint density function of y and z, as it can be obtained from $g(y^{s}, y^{d})$ by changing variables. We obtain the conditional density function of z, given y, as $h(y, z)/f(y)$. Then

$$\Pr\left(y^{d} \leq y^{s} | y\right) = \Pr(z \geq 0 | y)$$
$$= \int_{0}^{\infty} \frac{h(z, y)}{f(y)}\, \mathrm{d}z$$
$$= \frac{1}{f(y)} \int_{0}^{\infty} h(z, y)\, \mathrm{d}z.$$

To evaluate the right-hand-side integral we transform back to (y^s, y^d) again. As $z \geq 0$ corresponds to $y^d \leq y^s$ and, obviously, $h(z, y)$ transforms into $g(y^s, y^d)$, we end up with $f^{es}(y)$. This proves (A.4). Similarly we have

$$\Pr(y^s \leq y^d | y) = f^{ed}(y)/f(y). \tag{A.5}$$

It follows directly from (A.1) that both regime probabilities add to unity, as it should be.

We shall now demonstrate the *unboundedness* of the likelihood function L associated with a random sample y_i $(i = 1, \ldots, N)$ of observations on y:

$$L := \prod_{i \in I} f(y_i) = \prod_{i \in I} \{ f^{es}(y_i) + f^{ed}(y_i) \}, \tag{A.6}$$

where I is the index set $\{1, \ldots, N\}$. Without loss of generality we concentrate on finding a set of parameter values associated with the density function $g(y^s, y^d)$ such that L tends to infinity when the variance σ_d^2 of y^d tends to zero. This only occurs when we impose some, relatively weak, conditions on $g(y^s, y^d)$. As it is rather difficult, and not very rewarding, to try and find necessary conditions, we shall be content with the following set of sufficient conditions, where $g^s(y^s)$ and $g^d(y^d)$ are the marginal density functions of y^s and y^d, respectively:

Assumption. A set of admissible parameter values exists such that

$$g(y^s, y^d) = g^s(y^s) g^d(y^d), \; . \tag{A.7}$$

$$\forall i \in I, \qquad g^s(y_i) > 0, \tag{A.8}$$

$$\forall i \in I, \qquad y_i - \mathrm{E} y_i^d < 0, \tag{A.9}$$

$$\exists i' \in I, \qquad y_{i'} - \mathrm{E} y_{i'}^d = 0. \tag{A.10}$$

The first two conditions state the independence of y^s and y^d and the strict positivity of $g^s(\cdot)$ on the sample. In practice these conditions can always be met. The other two conditions are crucial in the proof. When, for instance, the model for y^d contains a constant term, i.e., $\mathrm{E} y_i^d = c + x_i'\beta$, where x_i is a vector of explanatory variables and β a vector of parameters, (A.9) and (A.10) can be satisfied for any value of β by choosing c equal to the maximum of $y_i - x_i'\beta$ over all observations. Without a constant term, however, no value of β may exist for which (A.9) can be satisfied, and the likelihood function will be bounded.

Now turning to the proof we choose a set of parameter values such that (A.7)–(A.10) are satisfied and let σ_d tend to zero. Substituting (A.2) and (A.3), and using (A.7) we obtain from (A.6),

$$L=\prod_{i\in I}\left[g^{\mathrm{d}}(y_i)\{1-G^{\mathrm{s}}(y_i)\}+g^{\mathrm{s}}(y_i)\{1-G^{\mathrm{d}}(y_i)\}\right], \tag{A.11}$$

where $G^{\mathrm{s}}(\cdot)$ and $G^{\mathrm{d}}(\cdot)$ are the distribution functions of y^{s} and y^{d}, respectively. When σ_{d} tends to zero the distribution of y^{d} degenerates. The density $g^{\mathrm{d}}(y)$ will tend to infinity for $y=\mathrm{E}y^{\mathrm{d}}$, and to zero elsewhere. The distribution $G^{\mathrm{d}}(y)$ tends to zero for $y<\mathrm{E}y^{\mathrm{d}}$, to one half for $y=\mathrm{E}y^{\mathrm{d}}$, and to unity for $y>\mathrm{E}y^{\mathrm{d}}$. Using these properties, and the fact that the distribution of y^{s} is not affected by σ_{d}, it can easily be checked from (A.8)–(A.10) that all factors in (A.11) will be strictly positive, whereas the factor corresponding to observation i' tends to infinity. This completes the proof.

When, finally, we introduce an *additional endogenous variable*, say an equation for prices p, we can reinterpret all results obtained so far as applying to the *conditional* distribution of y given p. Conditioning on p on both sides of (A.3), and multiplying with the marginal density function $g^{p}(p)$ of p, we obtain the joint density function $f'(y,p)$ of y and p as

$$\begin{aligned} f'(y,p)&=g^{p}(p)\{f^{\mathrm{es}}(y|p)+f^{\mathrm{ed}}(y|p)\}\\ &=f'^{\mathrm{es}}(y,p)+f'^{\mathrm{ed}}(y,p), \end{aligned} \tag{A.12}$$

where

$$f'^{\mathrm{es}}(y,p):=\int_{y}^{\infty}g'(y^{\mathrm{s}},y,p)\,\mathrm{d}y^{\mathrm{s}}, \tag{A.13}$$

$$f'^{\mathrm{ed}}(y,p):=\int_{y}^{\infty}g'(y,y^{\mathrm{d}},p)\,\mathrm{d}y^{\mathrm{d}}, \tag{A.14}$$

$g'(\cdot)$ being the joint density function of y^{s}, y^{d} and p. Conditioning on p on both sides of (A.4) and (A.5) we obtain the regime probabilities conditional on y and p as

$$\Pr(y^{\mathrm{s}}\le y^{\mathrm{d}}|y,p)=f^{\mathrm{es}}(Y|p)/f(y|p)=f'^{\mathrm{es}}(y,p)/f'(y,p), \tag{A.15}$$

$$\Pr(y^{\mathrm{d}}<y^{\mathrm{s}}|y,p)=1-\Pr(y^{\mathrm{s}}\le y^{\mathrm{d}}|y,p). \tag{A.16}$$

The likelihood function factorizes as

$$L=\prod_{i\in I}f'(y_i,p_i)=\prod_{i\in I}g^{p}(p_i)\prod_{i\in I}f(y_i|p_i).$$

Assuming strict positivity of $g^p(p_i)$ for all i, compare assumption (A.8), the unboundedness of L follows from the unboundedness of the second factor as before.

References

Fair, R.C. and D.M. Jaffee, 1972, Methods of estimation for markets in disequilibrium, Econometrica 40, 497–514.

Gersovitz, M., 1980, On classification probabilities for the disequilibrium model, Journal of Econometrics 14, 239–246.

Ginsburgh, V., A. Tishler and I. Zang, 1980, Alternative estimation methods for two-regime models, European Economic Review 13, 207–228.

Goldfeld, S.M. and R.E. Quandt, 1975, Estimation in a disequilibrium model and the value of information, Journal of Econometrics 3, 325–348.

Hartley, M.J. and P. Mallela, 1977, The asymptotic properties of a maximum likelihood estimator for a model of markets in disequilibrium, Econometrica 45, 1205–1220.

Kiefer, N.M., 1980, A note on regime classification in disequilibrium models, Review of Economic Studies 47, 637–639.

Kooiman, P., 1984, Smoothing the aggregate fix-price model and the use of business survey data, Economic Journal 94, 899–913.

Kooiman, P. and T. Kloek, 1979, Aggregation of micro markets in disequilibrium, Working paper (Econometric Institute, Erasmus University, Rotterdam).

Kooiman, P. and T. Kloek, 1985, An empirical two-market disequilibrium model for Dutch manufacturing, European Economic Review, forthcoming.

Laffont, J.-J. and R. Garcia, 1977, Disequilibrium econometrics for business loans, Econometrica 45, 1187–1204.

Lambert, J.-P., 1984, Disequilibrium macro-models based on business survey data: Theory and estimation for the Belgian manufacturing sector, Doctoral dissertation (CORE, Louvain-la-Neuve).

Lau, C.-L., 1980, A simple series for the incomplete gamma integral, Applied Statistics 29, 113–114.

Maddala, G.S., 1983, Limited-dependent and qualitative variables in econometrics (Cambridge University Press, Cambridge).

Maddala, G.S. and F.D. Nelson, 1974, Maximum likelihood methods for models of markets in disequilibrium, Econometrica 42, 1013–1030.

Malinvaud, E., 1982, An econometric model for macro-disequilibrium analysis, in: M. Hazewinkel and A.H.G. Rinnooy Kan, eds., Current developments in the interface: Economics, econometrics, mathematics (Reidel, Dordrecht).

Muellbauer, J., 1978, Macrotheory vs. macroeconometrics: The treatment of disequilibrium in macromodels, Discussion paper 59 (Birkbeck College, London).

Quandt, R.E., 1982, Econometric disequilibrium models, Econometric Reviews 1, 1–63.

Rosen, H.S. and R.E. Quandt, 1978, Estimation of a disequilibrium aggregate labour market, Review of Economics and Statistics 60, 371–379.

Thurik, A.R., 1984, Quantitative analysis of retail productivity (Meinema, Delft).

Thurik, A.R. and J. Koerts, 1984a, On the use of supermarket floorspace and its efficiency, in: The economics of distribution (Franco Angeli Editore, Milano).

Thurik, A.R. and J. Koerts, 1984b, Analysis of the use of retail floorspace, International Small Business Journal 2, 35–47.

Van Dijk, H.K., 1984, Posterior analysis of econometric models using Monte Carlo integration, Doctoral dissertation (Erasmus University, Rotterdam).

Van Dijk, H.K., T. Kloek and C.G.E. Boender, 1985, Posterior moments computed by mixed integration, Journal of Econometrics, this issue.