Tilburg University

Sampling for quality inspection and correction

Kleijnen, J.P.C.; Kriens, J.; Lafleur, M.

Publication date:
1986

Citation for published version (APA):
Kleijnen, J. P. C., Kriens, J., & Lafleur, M. (1986). *Sampling for quality inspection and correction: AOQL performance criteria*. (Research memorandum / Tilburg University, Department of Economics; Vol. FEW 246). Unknown Publisher.

UNIVERSITEIT

BRABANT

DEPARTMENT OF ECONOMICS

RESEARCH MEMORANDUM
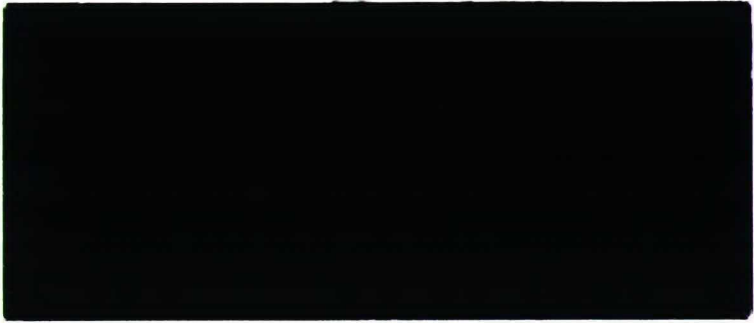
**SAMPLING FOR QUALITY INSPECTION AND
CORRECTION: AOQL PERFORMANCE CRITERIA**

J.P.C. Kleijnen
J. Kriens
M.C.H.M. Lafleur
J.H.F. Pardoel   S10.8

FEW 246

School of Business and Economics
Tilburg University
Post Box 90153
5000 LE Tilburg
Netherlands

SAMPLING FOR QUALITY INSPECTION AND CORRECTION:

AOQL PERFORMANCE CRITERIA

J.P.C. KLEIJNEN

J. KRIENS

M.C.H.M. LAFLEUR

J.H.F. PARDOEL

School of Business and Economics

Tilburg University

Post Box 90153

5000 LE Tilburg

Netherlands

ABSTRACT

The Average Outgoing Quality Limit (AOQL) denotes a sampling plan leading to inspection of the whole population, if the sample shows a number of defective items k exceeding an acceptance number k_o. According to the literature, this constant k_o is chosen such that the expected value of $\tilde{p}$, the fraction of defectives after inspection and possible correction, does not exceed a prespecified constant $\tilde{p}_m$; moreover several other performance criteria are estimated, using an extensive Monte Carlo simulation. The main conclusion is that the AOQL scheme is useful in practice, including applications in auditing, but the chance that the quality constraint is violated, $P\left[\tilde{\bar{p}} > \tilde{p}_m\right]$, may be sizable.

1. INTRODUCTION: AOQL

AOQL was introduced by Dodge and Romig around 1930; its application to auditing was studied by Kriens and Veenstra (1985). We can summarize this sampling scheme as follows. The goal of AOQL as introduced by Dodge and Romig, is to guarantee a minimum quality of the outgoing populations expressed as a maximum for the average fraction of defectives in the population. Kriens and Veenstra (1985) split up the yearly population into a number of subpopulations. The quality of a yearly population - before sampling and correction - is quantified by p, the percentage of "defective items" in the yearly population. The Number of items (defect plus correct) per Year is (say) NY, for example, a company produces NY cars; in auditing, accounts are sampled and NY is measured in dollars per years; see Kriens and Veenstra (1985, p. 387). Consequently, after inspection and correction the minimum quality corresponds to a maximum value for $\tilde{p}$.

The sampling scheme has the following steps (also see Table 1 later on).

(i) The expected yearly population is divided into a number of subperiods S, for example, S = 52 corresponds to production per week. These subpopulations may have different sizes, in expectation and certainly in realization. We denote the realized size of the subpopulation by N.

(ii) From each (realized) subpopulation a sample of size n is taken (n depends on several parameters, as we shall see).

(iii) Per sample (of size n) the number of defective items $\underline{k}$ is determined by inspection; obviously $\underline{k}$ is random (denoted by an underscore). And its integer values k satisfy: $0 \le k \le n$.

(iv) If and only if k exceeds a critical constant k_o (which varies with n; see step ii) then the whole subpopulation is

inspected and, by assumption, all defective items in the subpopulation are corrected. If $k \leq k_o$ then the defective items in the sample are corrected. So after sampling the quality of the subpopulation is improved, unless no defectives at all were found ($k = 0$) which probably indicates that the subpopulation had perfect quality already.

After step (iv) the fraction of defectives per subpopulation $\tilde{p}$ should satisfy the minimum-quality requirement $\tilde{p}_m$. So given a correct selection of the sampling plan's parameters n and k_o (see next paragraph), $\tilde{p}$ should satisfy the condition $E\left[\tilde{p}\right] \leq \tilde{p}_m$. Obviously, if the original fraction of defectives (before sampling) was very good already (say, $p = 0$), then $E\left[\tilde{p}\right] < p$. If this quality was very bad ($p \gg \tilde{p}_m$), then the sampling plan implies that sampling is (nearly) always followed by inspection and correction of the whole subpopulation so that ($0\approx$) $\tilde{p} \ll \tilde{p}_m$. See Figure 1 where p^* is the "least favorable" value of p.

Obviously $\underline{k}$ follows the hypergeometric distribution with parameters n, p and N. The critical constant k_o and n can be computed such that, not only holds $E\left[\tilde{p}\right] \leq \tilde{p}_m$, but also the expected costs are minimized. The original tables in Dodge and Romig (1959), however, contain some inaccuracies. Therefore we use the recent tables computed by Kriens and Winters (1987); see also Veenstra and Buysse (1985) and Van Batenburg, Kriens and Veenstra (1987). Table 1 illustrates some typical results.

In <u>practice</u> p, the before-sampling or prior probability, is unknown and often only the left-most columns in tables like Table 1 are used (low p). Even if the prior probability is estimated wrongly, the quality constraint $E\left[\tilde{p}\right] \leq \tilde{p}_m$ is met; the expected costs, however, may increase. A conjecture is that, not only does the expected value meet the quality constraint, but also the <u>chance</u> of too bad a yearly quality is <u>negligible</u>,
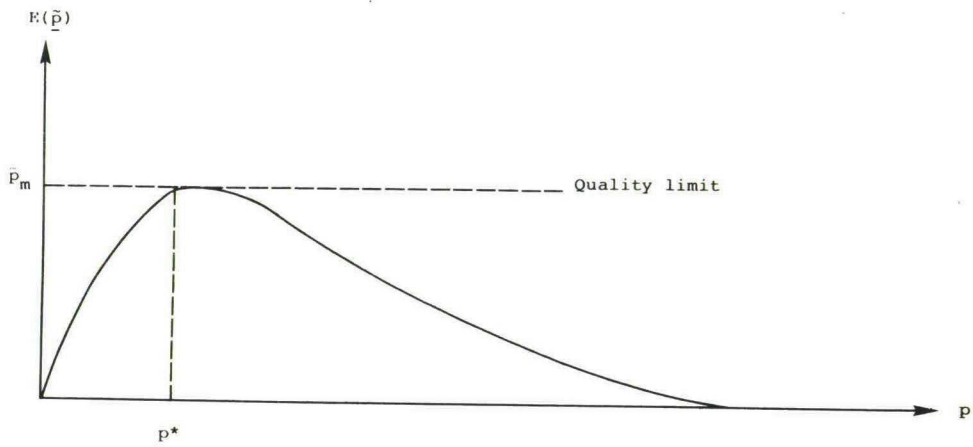
Figure 1: Expected fraction of defectives after sampling $E(\tilde{\underline{p}})$ versus fraction before sampling p.

Table 1: Sample size n and acceptance number k_o depending on prior probability p, subpopulation size N, and quality limit $\tilde{p}_m$ ($\tilde{p}_m$ = 1%).

Subpopulation size N	Before sampling p							
	0-0.02		...	0.21-0.40		...	0.81-1.00	
N	n	k_o		n	k_o		n	k_o
1-25	All	0		All	0		All	0
26-50	22	0		22	0		22	0
...								
801-1000	35	0		80	1		120	2
...								
1001-2000	36	0		80	1		180	3
20001-50000	85	1		255	4		990	15
50001-100000	85	1		255	4		1520	22

i.e., if $\bar{\tilde{p}}$ denotes the average yearly outgoing quality then

$$P\left[\bar{\tilde{p}} > \tilde{p}_m\right] \approx 0 \tag{1.1}$$

This conjecture is investigated in our simulation. Moreover, it hardly takes more computer time to estimate how bad $\bar{\tilde{p}}$ is _if_ the constraint $\bar{\tilde{p}} < \tilde{p}_m$ is violated, i.e., we estimate the following conditional expectations:

$$E\left[\bar{\tilde{p}} - \tilde{p}_m \,\middle|\, \bar{\tilde{p}} > \tilde{p}_m\right] \tag{1.2}$$

2. DESIGN OF MONTE CARLO EXPERIMENT

As Table 1 showed, the sample size n and acceptance number k_o are completely determined by the subpopulation size N, the prior probability p, and the quality limit $\tilde{p}_m$. In turn the subpopulation size N depends on the yearly population size NY and the number of subperiods S. In the simulation we study three values for S, namely 4, 13, and 52 which correspond to quarters, "months", and weeks. Our selection of the yearly population size NY is based in the experience of one of the authors with auditing applications: 10,000; 100,000 and 1,000,000. The expected subpopulation size $E(\underline{N})$ equals NY/S. We assume that the actual sizes $\underline{N}$ follow a uniform distribution with expected value NY/S; its range is such that the coefficient of variation is always (roughly) 6%, an arbitrarily selected value.

We selected the following six values for the quality limit $\tilde{p}_m$: 0.1%, 0.5%, 1%, 2%, 5%, 10%. Selection of the prior probability p in the simulation must be related to the quality limit $\tilde{p}_m$. There are no tables available for $p > 2 \tilde{p}_m$. This is no problem if only the left columns of the tables are used (see section 1); obviously if p is very high, then the scheme breaks down, that is, sampling is (nearly) always followed by inspection of the whole subpopulation; therefore we restrict our simulation of the "practitioner" to, $p \leq 6\ p_m$. Obviously not all subpopulations must have the same p, even if all subpopulations have the same expected value $E(\underline{p})$. Therefore we sample $\underline{p}$ in the simulation. As figure 1 demonstrates, the performance $E(\tilde{\underline{p}})$ improves as p deviates from the least favourable value p^*. In preliminary simulation experiments we sampled $\underline{p}$ from a distribution with a high variance, and indeed $E(\tilde{\underline{p}})$ decreased (not further reported in this paper). Therefore we concentrate our simulation on worst case situations, that is, $\underline{p}$ has a small

range. We further assume that $\underline{p}$ is uniformly distributed with a range of only $0.2\ \tilde{p}_m$. We do change the expected value $E[\underline{p}]$; as we explained above, we vary p between 0 and $6\ \tilde{p}_m$. So we sample $\underline{p}$ from the uniform distribution between 0 and $0.2\ \tilde{p}_m$, between $0.2\ \tilde{p}_m$ and $0.4\ \tilde{p}_m$, ... , between $5.8\ \tilde{p}_m$ and $6\ \tilde{p}_m$.

Summarizing, we simulate 1620 factor combinations using only the left-most columns of the tables ("practitioner's approach") and 540 combinations with the optimal $[n,\ k_o]$ combinations ("theoretical approach").

There is a technical issue in the simulation: how often should each factor combination be simulated in order to obtain <u>reliable estimates</u> of performance criteria such as $P\left[\bar{\tilde{p}} > p_m\right]$? By definition, one replicate yields a binomial variable (say) $\underline{x}$ with $q = P(\underline{x} = 0) = P\left[\bar{\tilde{p}} > \tilde{p}_m\right]$. Using the normal approximation to the binomial distribution, it is straightforward to derive the number of replications needed to estimate q with either a relative precision of 10% or an absolute precision of 0.001; also see Kleijnen (1987, pp. 46-51). This approach shows that at most 16,221 replications are needed (when $\hat{q}$ = .01) to satisfy either the relative precision or the absolute precision requirement, with 90% probability. Actually we do not know q. So we substitute the "current" estimate for q (after at least 100 replications) that is, the estimate available after (say) r replications where r = 101, 102, ... We found that the average number of replications is roughly 1000. We emphasize that the simulation not only estimates the performance criterion q = $P\left[\bar{\tilde{p}} > \tilde{p}_m\right]$ but several more criteria. The main criterion, however, is q so that we concentrate on q to select the number of replications.

It turns out that it takes 40 hours of computer time to simulate 1620 + 540 factor combinations, each combination re- plicated roughly 1000 times. To keep computer time below this (sizable) value, we have to introduce the following technical refinement. The number of defectives $\underline{k}$ has a hypergeometric distribution. The binomial distribution provides a good ap- proximation of $n \ll N$ which is often the case (but not always: if NY is small then $n > N$ may occur); see Table 1. The Poisson distribution is a good approximation to the binominal distribu- tion, if p is small. We simulate the Poisson distribution using the subroutine in Naylor et al. (1966, p. 114), so that our program runs 20 times faster on our computer (a VAX 780 running under VMS using the NAG subroutine for the multiplicative con- gruential random number generator).

3. MONTE CARLO RESULTS

The Monte Carlo experiment yields a mass of data. We analyze these data through regression analysis (using SAS). Preliminary plots look like a gamma function. Therefore we fit such a type of non-linear regression model for $\bar{\tilde{p}}$; its R^2 adjus- ted for the number of explanatory variables, is higher than 0.95 and yields Figure 2, which looks like the theoretical Figure 1: $\bar{\tilde{p}} \leq \tilde{p}_m$.

If the prior probability satisfies $p \leq \tilde{p}_m$ then obvious- ly $q = P\left[\bar{\tilde{p}} > \tilde{p}_m\right] = 0$. If $p > \tilde{p}_m$ then we again fit a function like the gamma function, with $R^2 = 0.99$ for the theoretical case and 0.74 for the practitioner's approach; see Figure 3. So there is a sizable chance (up to 40% in Figure 3) of violating the quality constraint $\bar{\tilde{p}} \leq \tilde{p}_m$. However, we repeat that our simulation is a worst case, since p of the subpopulation is sampled from a uniform distribution with a small range.
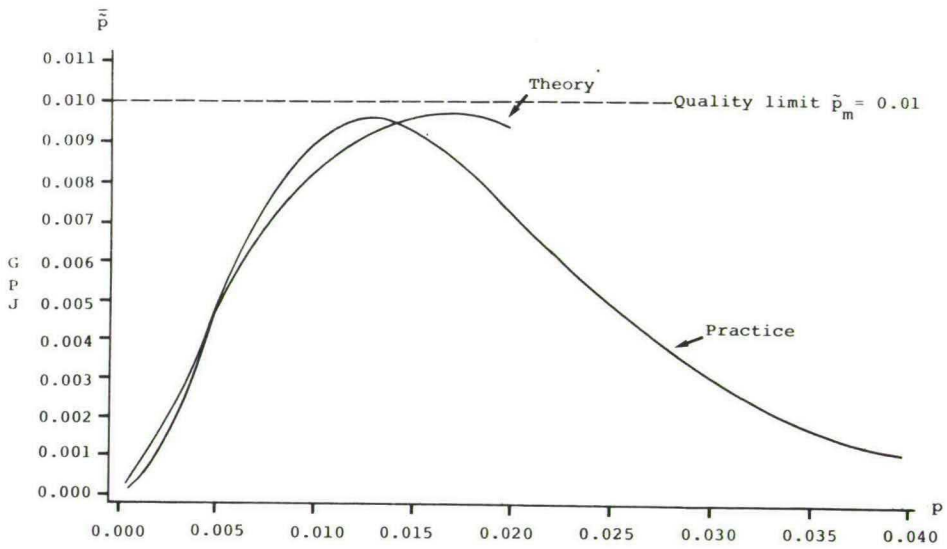
Figure 2: Average yearly outgoing quality $\bar{\bar{P}}$ versus fraction of defectives before sampling p (yearly population NY = 1,000,000; subperiods S = 52).
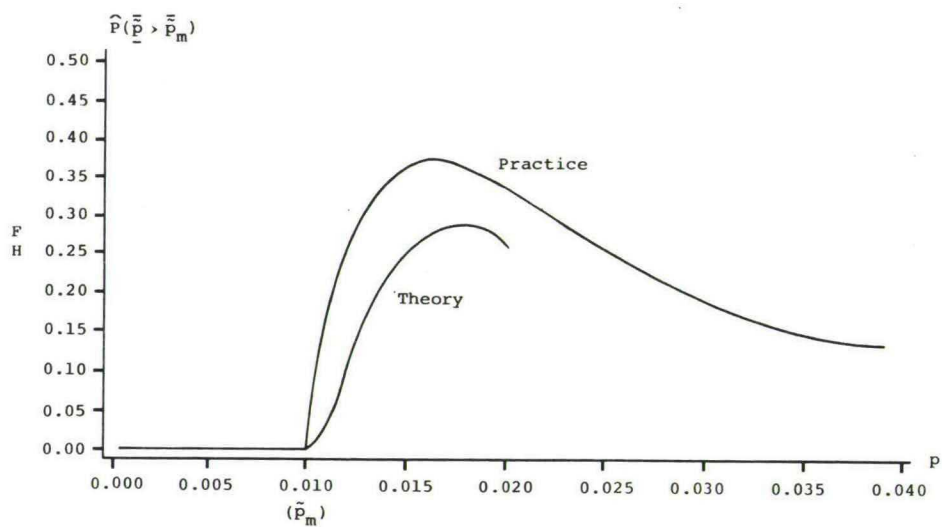
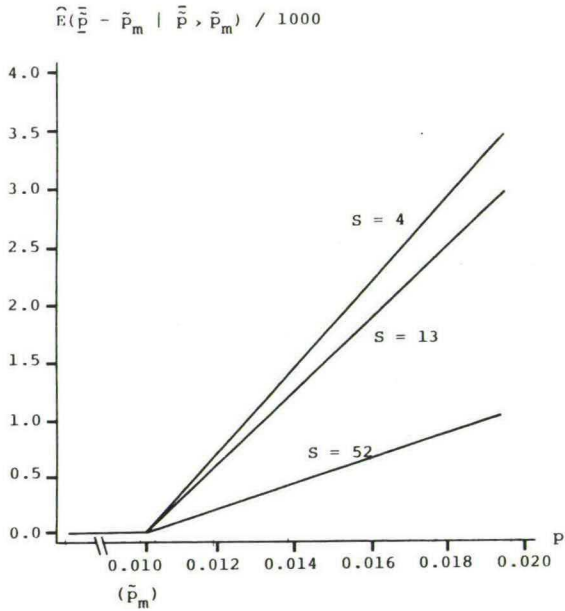Figure 3: Estimated probability of quality violation, $P(\bar{\bar{p}} > \tilde{p}_m)$ (NY = 100,000; S = 52).

Figure 4: Estimated size of quality violation, $E(\bar{\bar{p}} - \tilde{p}_m | \bar{\bar{p}} > \tilde{p}_m)$ in theoretical approach (NY = 1,000,000).
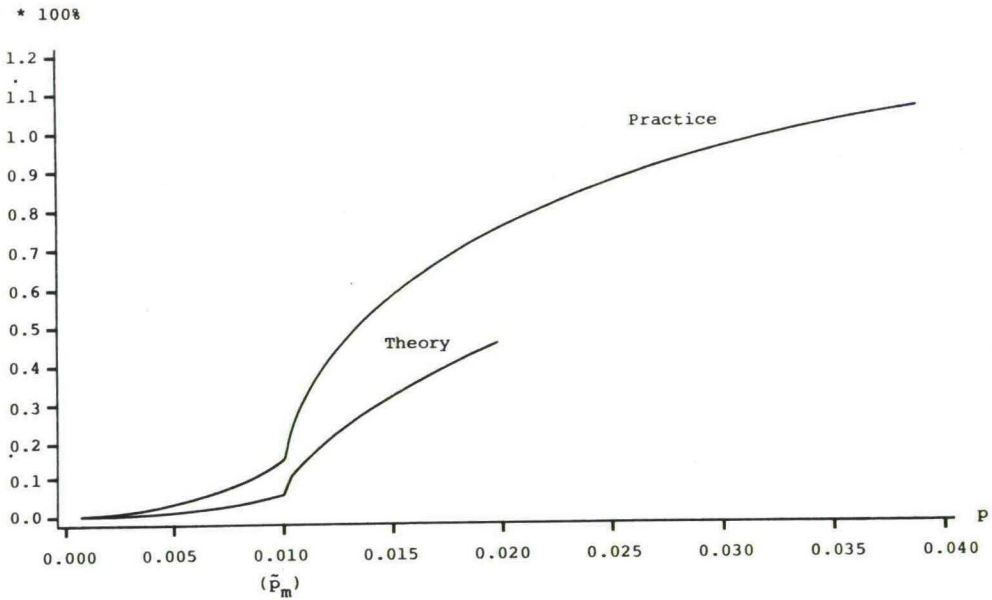
Figure 5: Estimated fraction of yearly population, actually inspected
(NY = 1,000,000 ; S = 52).

If $\bar{\tilde{p}} > \tilde{p}_m$ then we wonder how bad the quality violation is: $E\left[\bar{\tilde{p}} - \tilde{p}_m | \bar{\tilde{p}} > \tilde{p}_m\right]$. Figure 4 shows that smaller subperiods (higher S) give extra protection.

Next we consider the costs of the AOQL scheme. This scheme implies that all N units (of a subperiod) are inspected if $k > k_o$. Figure 5 shows the fraction of the yearly population actually inspected. That fraction increases drastically if $p > \tilde{p}_m$. Obviously the practitioner's approach is more expensive. We add that the curves are hardly effected by S, the number of subperiods (not displayed).

Our simulation shows that is it important to have a good idea about p, the before sampling fraction of defectives. Therefore we suggest to obtain an estimate of p, using $\hat{p} = k/n$ if $k \leq k_o$ and $\hat{p} = \underline{K}/\underline{N}$ if $k > k_o$ where $\underline{K}$ denotes the number of defectives in the subpopulation (of size $\underline{N}$). As time t goes on, the estimators $\hat{p}_t$ can be combined; for example, we may weigh the $\hat{p}_t$ with the sample sizes n_t or the subpopulation sizes N_t (if $k \leq k_o$ or $k > k_o$ respectively). If $\hat{p}_t$ shows serial correlation or non-stationary behavior, then time series techniques may be applied.

4. CONCLUSIONS

The AOQL sampling plan is indeed used in practice; see Kriens and Veenstra (1985). Then it is assumed that if the expected yearly fraction of defectives after inspection and correction $E(\tilde{p})$ meets the quality constraint $\tilde{p}_m$ then the probability of exceeding the constraint, $P\left[\bar{\tilde{p}} > \tilde{p}_m\right]$ is negligible. Figure 3 (based on our simulation data analyzed by regression) shows that this probability is sizable, if the "before inspection" or prior probability p is higher than the limit $\tilde{p}_m$ but

not extremely high (if $p \gg \tilde{p}_m$ then most times sampling is followed by inspection of the whole subpopulation). If in practice p varies much over subperiods, then $P\left[\bar{\tilde{p}} > \tilde{p}_m\right]$ decreases (we simulated worst case situations: small range of $\underline{p}$). One can get an estimate of p from the sampling procedure: if $k \leq k_o$ then $\hat{p} = \underline{k}/n$; else $\hat{p} = \underline{K}/\underline{N}$. Increasing the number of periods S decreases the size of the expected quality violation; see Figure 4. Underestimating p is not wise: it does not give extra quality protection (Figure 3) and yet more inspection work is done (Figure 5). So in practice one should build up knowledge about p.

REFERENCES

Dodge, H.F. and H.G. Romig, *Sampling inspection tables*. John Wiley, New York, second edition, 1959.

Kleijnen, J.P.C., *Statistical tools for simulation practitioners*. Marcel Dekker, Inc., New York, 1987.

Kriens, J. and R.H. Veenstra, Statistical sampling in internal control by using the AOQL-system. *The Statistician*, 34, 1985, pp. 383-390.

Kriens, J. and A.A.G.M. Winters (1987), The A.O.Q.L. tables by Dodge and Romig: some corrections. (to appear.)

Naylor, T.H., J.L. Balintfy, D.S. Burdick and K. Chu, *Computer simulation techniques*. John Wiley & Sons, Inc., New York, 1966.

Van Batenburg, P.C., J. Kriens and R.H. Veenstra (1987), Average Outgoing Quality Limit - a revised and improved version. (to appear in Dutch.)

Veenstra, R.H. and J.C. Buysse (1985), Optimizing sampling applications in the administration, *De Accountant*, 10, no. 198, pp. 561-563. (in Dutch.)

IN 1986 REEDS VERSCHENEN

202 J.H.F. Schilderinck
     Interregional Structure of the European Community. Part III

203 Antoon van den Elzen and Dolf Talman
     A new strategy-adjustment process for computing a Nash equilibrium in
     a noncooperative more-person game

204 Jan Vingerhoets
     Fabrication of copper and copper semis in developing countries. A
     review of evidence and opportunities

205 R. Heuts, J. van Lieshout, K. Baken
     An inventory model: what is the influence of the shape of the lead
     time demand distribution?

206 A. van Soest, P. Kooreman
     A Microeconometric Analysis of Vacation Behavior

207 F. Boekema, A. Nagelkerke
     Labour Relations, Networks, Job-creation and Regional Development. A
     view to the consequences of technological change

208 R. Alessie, A. Kapteyn
     Habit Formation and Interdependent Preferences in the Almost Ideal
     Demand System

209 T. Wansbeek, A. Kapteyn
     Estimation of the error components model with incomplete panels

210 A.L. Hempenius
     The relation between dividends and profits

211 J. Kriens, J.Th. van Lieshout
     A generalisation and some properties of Markowitz' portfolio selecti-
     on method

212 Jack P.C. Kleijnen and Charles R. Standridge
     Experimental design and regression analysis in simulation: an FMS
     case study

213 T.M. Doup, A.H. van den Elzen and A.J.J. Talman
     Simplicial algorithms for solving the non-linear complementarity
     problem on the simplotope

214 A.J.W. van de Gevel
     The theory of wage differentials: a correction

215 J.P.C. Kleijnen, W. van Groenendaal
     Regression analysis of factorial designs with sequential replication

216 T.E. Nijman and F.C. Palm
     Consistent estimation of rational expectations models

IN 1987 REEDS VERSCHENEN

242   Gerard van den Berg
      Nonstationarity in job search theory

243   Annie Cuyt, Brigitte Verdonk
      Block-tridiagonal linear systems and branched continued fractions

244   J.C. de Vos, W. Vervaat
      Local Times of Bernoulli Walk

245   Arie Kapteyn, Peter Kooreman, Rob Willemse
      Some methodological issues in the implementation
      of subjective poverty definitions