Optimal design of simulation experiments with nearly saturated queues

Cheng, R.C.H.; Kleijnen, J.P.C.

Publication date:
1995

entER
for
omic Research

Discussion paper

[redacted]

Tilburg University

[redacted]

Center
for
Economic Research

No. 9567

OPTIMAL DESIGN OF SIMULATION EXPERIMENTS WITH NEARLY SATURATED QUEUES

By Russell C.H. Cheng
and Jack P.C. Kleijnen

June 1995

ISSN 0924-7815

Optimal Design of Simulation Experiments with Nearly Saturated Queues

Russell C. H. Cheng

Institute of Mathematics and Statistics, University of Kent at Canterbury, Canterbury, Kent, CT2 7NF, United Kingdom.

Fax: +44-1227-475473. E-mail: R.C.H.Cheng@ukc.ac.uk

Jack P. C. Kleijnen

CentER/Department of Information Systems and Auditing, School of Management and Economics, Tilburg University, Tilburg, 5000 LE Tilburg, Netherlands

Fax: +31-13-663377. E-mail: kleijnen@kub.nl

January 1995

Optimal Design of Simulation Experiments with Nearly Saturated Queues

Russell C. H. Cheng

Institute of Mathematics and Statistics, University of Kent at Canterbury, Canterbury, Kent, CT2 7NF, United Kingdom.

Jack P. C. Kleijnen

CentER/Department of Information Systems and Auditing, School of Management and Economics, Tilburg University, Tilburg, 5000 LE Tilburg, Netherlands

Abstract

Regression analysis can be used to approximate the input/output behaviour of simulation models; this yields 'metamodels'. This paper focuses on simulation models with heavy traffic. The regression model is assumed to be linear in its parameters but not in its independent variables. The simulation responses are supposed to have variances that increase with the traffic load of the simulated queueing system. Computer time is supposed to be limited. Two main questions are addressed: which input values to simulate, and how many customers to simulate at each of these input values? Terminating simulations are distinguished from steady-state simulations. Typically,

the highest traffic load actually simulated should remain far away from its upper limit; but a large percentage of the available computer time should be allocated to simulating this highest rate. Simulation examples are included, demonstrating that optimal designs can improve the accuracy of the estimated metamodel considerably.

Keywords: response surface, interpolation, extrapolation, runlength

Introduction

Recently the analysis and design of simulation experiments with models of nearly saturated queues has received considerable attention. Whitt (1989) and Asmussen (1992) discuss the behaviour of steady-state queues as the saturation limit is approached; Whitt in particular considers how the length of simulation runs should be increased in order to maintain accuracy in estimating queue length or waiting time. In this paper we consider the behaviour of queues over a wide range of traffic intensities including values approaching saturation. We examine both steady-state and transient behaviour.

A natural way of examining the response over a range of input values is, as suggested by Kleijnen (1975a, 1992), to use a *regression* model as a metamodel of the input-output behaviour of the underlying simulation model. Cheng (1990) also considers such models, and shows how conditional sampling can be incorporated as a variance reduction method to improve accuracy. Reiman, Simon and Willie (1992) consider regression models focusing on how known theoretical results about light traffic and heavy traffic behaviour can be incorporated into the analysis, to improve accuracy. Cheng and Traylor (1993) show how conditional sampling and use of known theoretical results can be combined. Vollebregt (1994a,b) investigates a problem similar to our problem. He, however, reverses the roles of the criterion and the budget condition (he minimises computer time, given a certain statistical accuracy); he does not concentrate on heavy traffic queueing.

An aspect that has not been addressed seriously in the literature is how to optimally design the overall simulation experiment in order to minimize the variances of estimators of interest. Optimal design in regression analysis was discussed in the seminal paper, Kiefer and Wolfowitz (1959), to whom Reiman et al. (1992) also refer. However, Kiefer and Wolfowitz focus on the case where the response variable has constant variance. A characteristic feature of nearly saturated queues is that the variability of the response of interest, such as waiting time, becomes unbounded as the saturation limit is approached. The Kiefer and Wolfowitz approach can in principle be adjusted to handle this case. However, it turns out that the modifications are quite involved; moreover, heteroscedasticity gives solutions that differ substantially from the homoscedastic case. We show instead that, for a certain class of regression models, the problem can be solved directly using a result (Theorem 2; see eqs. 16 and 17) which allows us to explicitly handle heteroscedasticity. The theorem shows how the variance of the response depends on the values of the independent (input) variable and on the number of replications used at each input value. This characterizes the variance, and makes its minimization straightforward. We give examples showing that the typical optimal design has, as its highest traffic load, a value that is far away from the saturation limit; but that a large percentage of the available computer time is allocated to simulating at this highest rate.

This paper is organised as follows. In Section 2 we introduce the regression model to be considered. In Section 3 we give the theorem and show its use in the optimal design of simulation experiments. In Section 4 we study the amount of computing effort needed to make a simulation run. This obviously needs to be taken into account in assessing overall efficiency, and is an important aspect that is not always adequately discussed. We show that a clear distinction needs to be made between steady-state simulations and terminating simulations, and show how the optimal design is modified accordingly. In Section 5 we give examples involving the M/M/1 queue and compare our analytical results with simulation results. These examples show good agreement between the two, and illustrate the substantial efficiency gains possible with our approach. In Section 6 we give conclusions, and discuss side issues and extensions, including additional references.

1 The Regression Metamodel

We suppose that the simulation experiment is made up of a number of independent runs. We assume that y, the output (response) of a run, is determined by x, an independent input variable, and that this input/output relationship can be represented by the following regression metamodel (response surface):

$$y_{ij} = (\gamma_0 + \gamma_1 x_i + \gamma_2 x_i^2 + ... + \gamma_k x_i^k) f(x_i) + \epsilon_{ij} \quad (i = 1, ..., n) \quad (j = 1, ..., m_i) \tag{1}$$

with

ϵ: approximation error of the metamodel, with mean 0 and variance σ_i^2;

$\gamma = (\gamma_1, ..., \gamma_k)'$: vector of k unknown parameters representing input effects;

$f(x)$: a known function (discussed in the next paragraph).

We make the simulation runs at only n distinct input values $x_1, x_2, ..., x_n$, with m_i observations placed at the ith point, x_i $(i = 1, ..., n)$. We denote the total number of runs by N; that is, $\sum_{i=1}^{n} m_i = N$.

The purpose of introducing the function $f(x)$ is to allow regression models where the response can become unbounded. In particular it allows us to consider saturated queueing situations. Consider for example the M/M/1 queue. If the arrival rate is unity, then the steady-state expected waiting time has the form

$$E(y) = x/(1 - x) \tag{2}$$

where x is the traffic intensity. If we do not know the correct expression for $E(y)$ but know only that queue saturation occurs as $x \to 1$, then we may assume that the responses from the simulation runs have the form

$$y_i = (\gamma_0 + \gamma_1 x_i + \gamma_2 x_i^2)/(1 - x_i) + \epsilon_i \tag{3}$$

where the vector of coefficients $\gamma = (\gamma_0, \gamma_1, \gamma_2)'$ is unknown and is to be estimated. This model is of the form (1). We shall return to this example.

Another example is the problem discussed by Reiman et al. (1992) where the expected squared waiting time $E(y^2)$ is to be estimated in an M/G/1 queue where the service time has a certain mixture distribution. In their example it turns out that

$$E(y^2) = (\frac{5}{2} + x + \frac{3}{8}x^2)/(1-x)^2 \tag{4}$$

where x is again the traffic intensity. In this case $f(x) = (1-x)^{-2}$ and we see that (4) is correctly represented by a regression of the form (1).

In practice the precise form of $f(x)$ may not be known, or only be known approximately. When available, heavy traffic queueing theory can be used to suggest of an appropriate form for $f(x)$. The polynomial component is therefore included in (1) to provide flexibility in the metamodel to compensate for any approximation in $f(x)$.

In this paper we do not assume that the error variance is constant; it may depend on i, or equivalently, on x_i. For example, Whitt (1989) and Asmussen (1992) show that the variance of ϵ in (3) is $O[(1-x)^{-4}]$ as $x \to 1$. As pointed out by Kiefer and Wolfowitz (1959), if the form of the dependence is known, then homogeneity of variance can be restored as follows. Let

$$Var(\epsilon) = [v(x)\sigma]^2 \tag{5}$$

where $v(x)$ is known. Then we simply divide (1) by $v(x)$ and get

$$\begin{aligned} z_{ij} &= y_{ij}/v(x_i) \\ &= (\gamma_0 + \gamma_1 x_i + \gamma_2 x_i^2 + \ldots + \gamma_k x_i^k) r(x_i) + \delta_{ij} \end{aligned} \tag{6}$$

where

$$r(x_i) = f(x_i)/v(x_i) \tag{7}$$

and $Var(\delta_{ij}) = \sigma^2$. In the example (3) Whitt and Asmussen's result means multiplying y by $(1-x)^2$, so (3) becomes

$$z_i = (\gamma_0 + \gamma_1 x_i + \gamma_2 x_i^2)(1 - x_i) + \delta_i \tag{8}$$

where $Var(\delta)$ is now independent of x.

We assume that the objective of the simulation is to estimate some known linear function of the coefficients $\gamma_0, \gamma_1, ..., \gamma_k$ in (1):

$$p = \alpha_0\gamma_0 + ... + \alpha_k\gamma_k = \alpha'\gamma \tag{9}$$

where the vector $\alpha = (\alpha_0, \alpha_1, ..., \alpha_k)'$ is known. In the example (3) we might be particularly interested in the behaviour of the queue as $x \to 1$. Thus we would wish to estimate

$$p = \lim_{x \to 1}(\gamma_0 + \gamma_1 x + \gamma_2 x^2) = \gamma_0 + \gamma_1 + \gamma_2. \tag{10}$$

We use the Ordinary Least Squares (OLS) estimator, after having applied the variance stabilizing transformation (OLS is identical to Weighted Least Squares applied to the original observations $\mathbf{y}$):

$$\hat{\gamma} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{z} \tag{11}$$

where $\mathbf{z} = (z_{11}, z_{12}, ..., z_{nm_n})'$ is the vector of simulation responses (of dimension N), and

$$\mathbf{X} = \begin{pmatrix} \mathbf{x_1} \\ \cdots \\ \mathbf{x_2} \\ \cdots \\ \cdots \\ \mathbf{x_n} \\ \cdots \end{pmatrix}$$

is the $N \times (k+1)$ matrix of independent variables, with the row vector

$$\mathbf{x_i} = (1, x_i, x_i^2, ..., x_i^k) r(x_i)$$

appearing m_i times in $\mathbf{X}$. The matrix $\mathbf{X'X}$ is persymmetric (that is, all elements in any diagonal perpendicular to the main diagonal are all the same; see, for example, Aitken, 1964, p.121), its i,jth entry being

$$(\mathbf{X'X})_{ij} = \sum_{l=1}^{n} n_l x_l^{i+j} \quad (i, j = 1, 2, ..., k) \tag{12}$$

where $n_i = m_i r^2(x_i)$. for $i = 1, 2, ..., n_i$. We estimate p in (9) using

$$\hat{p} = \alpha_0 \hat{\gamma}_0 + ... + \alpha_k \hat{\gamma}_k. \tag{13}$$

Our main objective is to show how to select the m_i and the x_i in order to minimize the variance of (13). We shall discuss this in the next section.

2 Optimal Selection of Simulation Inputs

Kiefer and Wolfowitz (1959) address a general version of the optimum design problem presented in Section 2. They point out that, in general, a direct approach to the problem leads rapidly to intractable algebra. They give an ingenious approach that transfers the difficulty into a Chebyshev approximation problem, thus allowing use of the extensive literature on this topic. They give examples drawn from polynomial regression.

It is in principle possible to make use of the Kiefer and Wolfowitz approach in our problem. However, there are two aspects that make an alternative, direct, approach competitive in our case. Firstly, it turns out that the persymmetric form of the matrix $\mathbf{X}'\mathbf{X}$ makes the direct approach much more tractable. Secondly, the form of our regression function, being non-standard, gives rise to a non-standard Chebyshev approximation that is arguably no easier, and is possibly more difficult, to solve than the problem resulting from our direct approach.

We start with the optimal choice of n (the number of distinct x values). Kiefer and Wolfowitz (1959, Theorem 2) show that if the functions multiplying the coefficients are linearly independent, then n should be chosen equal to the number of unknowns, $k+1$ in our case. Also see Vollebregt (1994b, p.13). We assume $n = k+1$ from now on.

We now investigate how to choose the m_i (the number of replications at x_i, with $i = 1, 2, ..., n$).

Theorem 1: Let $\mathbf{A} = \mathbf{X}'\mathbf{X}$ be as defined in (12). Then the inverse of $\mathbf{A}$ has the form:

$$\mathbf{A}^{-1} = \sum_{i=1}^{k+1} \mathbf{B}_i / [m_i r^2(x_i)] \tag{14}$$

where the matrices $\mathbf{B}_i$ are independent of the m_i and the form of $r(.)$, and are positive semi-definite.

Proof: See Appendix.

Corollary: The variance of $\hat{p}$, as given in (13), has the form

$$Var(\hat{p}) = \sigma^2 \sum_{i=1}^{k+1} \left(\frac{a_i}{r_i}\right)^2 m_i^{-1} \tag{15}$$

where $a_i = \sqrt{\alpha' \mathbf{B}_i \alpha}$, $i = 1, 2, ..., k+1$, and for simplicity we have written r_i for $r(x_i)$.

Theorem 1 allows us to explicitly calculate the optimal m_i that will minimize (15), given x_i. We have the following theorem.

Theorem 2: Let the simulation points $x_1, x_2, \ldots, x_{k+1}$ be fixed, and let the total number of simulations be equal to N, so that $\sum_{i=1}^{k+1} m_i = N$. Then the variance (15) is minimized if

$$m_i = \frac{a_i/r_i}{(\sum_{j=1}^{k+1} a_j/r_j)} N \quad (i = 1, 2, ..., k+1). \tag{16}$$

The minimized value is

$$V_{\min}(x_1, ..., x_{k+1}) = \frac{\sigma^2}{N} [\sum_{j=1}^{k+1} (a_j/r_j)]^2. \tag{17}$$

Proof: Minimization of (15) subject to $m_i > 0$, all i, and $\sum_{i=1}^{k+1} m_i = N$ is a convex programming problem. An easy application of the method of Lagrange multipliers gives the result. □

In simple cases, Theorem 2 is sufficient to enable the optimal settings of the x_i to be found explicitly by solving the system of equations

$$\partial V_{\min}(x_1, ..., x_{k+1}) / \partial x_i = 0 \quad (i = 1, 2, ..., k+1).$$

Though it is possible to write down an explicit formula for the $\mathbf{B}_i$ of eq. (14), and hence a_i, it is usually easier to identify the ratios a_i/r_i needed in (16) and

(17) by manipulating $Var(\hat{p})$ into the form (15), and obtaining a_i/r_i from this expression directly. In more complicated cases it is probably easiest to numerically minimize (15) using a standard algorithm. We give examples in Section 5.

3 Translating Regression Variables into Simulation Variables

The definition of a simulation observation needs to be done with some care. To simulate one customer, the analyst must generate one interarrival and one service time (in more complicated queueing systems the analyst must sample more random variables; we concentrate on GI/G/1). Without loss of generality we suppose that it takes one unit of computer time to generate and process one customer. It is convenient to measure run length in such units, as the cost involved is then directly measured in run length and there is no ambiguity whether we talk of the number of customers processed or of run length.

Kleijnen (1974/1975) distinguishes between *terminating* simulations and *steady-state* simulations. We shall discuss the latter first, as they are actually easier to analyse in our terms, even though they are harder to carry out in practice.

3.1 Steady-State Simulations

Suppose that a total of C units of computer time is available and that at the simulation point x_i the runs are of length b_i. Then we have the new constraint $\sum b_i m_i = C$, as well as the original constraint $\sum m_i = N$. In the case of steady-state simulation it is well known that, if the response is some form of sample average, then its variance is $O(b_i^{-1})$ as $b_i \to \infty$. Thus, provided the b_i are sufficiently large, we can write (5) in the form

$$Var(\epsilon_{ij}) \simeq [w(x_i)\sigma]^2/b_i \tag{18}$$

where $w(x_i)$ does not depend on m_i or b_i. Then (5), (7) and (18) yield $r(x_i) = f(x_i)b_i^{\frac{1}{2}}/w(x_i)$. Defining $s_i = f(x_i)/w(x_i)$, (15) becomes

$$Var(\hat{p}) = \sigma^2 \sum_{i=1}^{k+1} \left(\frac{a_i}{s_i}\right)^2 (b_i m_i)^{-1}. \tag{19}$$

We see therefore that minimization of (19) subject to $\sum b_i m_i = C$, is precisely the same problem as the original, except that $b_i m_i$ replaces m_i, and s_i replaces r_i. The solution is therefore analogous to (16):

$$b_i m_i = \frac{a_i/s_i}{(\sum_{j=1}^{k+1} a_j/s_j)} C \quad (i = 1, 2, ..., k+1). \tag{20}$$

Thus we have the result that the only requirement is that $b_i m_i$, the total run length at the point x_i, should always be as given by (20). We are free to choose either the value of b_i (that is, how long to make each run), or to choose the value of m_i (that is, how many runs to make at the point x_i). Whichever value is chosen, the other is given by (20). Consequently the total number of observations N can be regarded as either fixed or variable, as we like. For example, if N is given, we can choose any set of m_i satisfying $\sum m_i = N$, and then fix the run lengths using (20). The only proviso is that m_i should not be so large that it makes b_i too small, because (18), which is an asymptotic result, might then no longer hold.

We can now consider different choices for the simulation run length. To be explicit we consider the example of the M/M/1 queue where we wish to estimate the expected waiting time. Whitt (1989) has shown that as the traffic intensity x tends to unity, the variance of the estimated waiting time (from a run of given length) is $O[(1-x)^{-4}]$. In our formulation this means we can set $w(x) = (1-x)^{-2}$.

If we wish to have run lengths of equal length, we simply simulate the same number of steady-state customers in each run: $b_i = b$ for all i. However, this will mean that the responses from different runs will have different variances, as given by (18).

If we want runs of equal variance, then run lengths should be proportional to $(1-x)^{-4}$; that is, we set $b_i = b(1-x)^{-4}$.

Another possibility is to use *renewal* (or regenerative) analysis and use a cycle as the basic run length. In the queue example the expected length of a cycle is $x/(1-x)$. In this case we therefore would set $b_i = bx/(1-x)$. Also see Asmussen (1992).

The number of runs N is in effect arbitrary, so we could in principle set $N = k+1$, and make a single run at each point: $m_i = 1$ for all i. However, we would then need to use batching or spectral analysis, say, to assess lack of fit. Conversely, making the m_i too large would mean making very many short runs. In general this is a bad idea, as each run will require a setup time to reach steady state, and it is usually more efficient to make one long run (see Cheng, 1976).

3.2 Terminating Simulations

A terminating queueing simulation is usually concerned with estimating transient characteristics of a queue. Often the arrival and service rates will not be constant, and the traffic intensity will then be a function of time. It will usually therefore not be appropriate to consider the simulation input, x, as a specific traffic intensity; instead x is some more generalised measure of queue activity level.

Moreover, unlike the steady-state case, the definition of what constitutes a run will be predefined by the specification of the terminating event, so its length is not a quantity we can arbitrarily select. We therefore write the run length as $b(x_i)$ to indicate its dependence on the simulation input. The error variance is still given by (5), and $Var(\hat{p})$ by (15). The optimum choice of m_i is obtained by minimizing $Var(\hat{p})$ subject to $\sum b(x_i)m_i = C$. This gives

$$m_i = \frac{a_i/[r_i b^{\frac{1}{2}}(x_i)]}{\{\sum_{j=1}^{k+1} a_j/[r_j b^{\frac{1}{2}}(x_j)]\}} C \quad (i = 1, 2, ..., k+1). \tag{21}$$

Obviously in this case we cannot select the total number of runs, N, independently of C.

4 Examples

4.1 Steady-State Waiting Times in the M/M/1 Queue

We consider the example based on the waiting time response curve (3). In this case, $f(x)$ in (1) becomes $(1-x)^{-1}$. If we assume that simulation runs are of equal length (we simulate the same number of customers), then $Var(\epsilon) \simeq (1-x)^{-4}\sigma^2$ as $x \to 1$. Consequently, in (5) we have $v(x) = (1-x)^{-2}$, and in

(7) we get $r(x) = (1 - x)$. Strictly speaking, we need to have an expression for $Var(\epsilon)$ that is valid for the entire range of interest of x. However, as we are especially interested in the behaviour of the queue as $x \to 1$, we shall apply this limiting expression for $Var(\epsilon)$, for all x, when selecting our design.

Our objective is to estimate the sensitivity of the queue as $x \to 1$, so we wish to minimize the variance of the estimator of (10). After some algebra, we find that this variance has the form

$$V(\hat{p}) = (c_1^2/m_1 + c_2^2/m_2 + c_3^2/m_3)\sigma^2$$

where

$$c_1 = \frac{(1-x_2)(1-x_3)}{(x_2-x_1)(x_3-x_1)(1-x_1)},$$
$$c_2 = \frac{(1-x_1)(1-x_3)}{(x_2-x_1)(x_3-x_2)(1-x_2)},$$
$$c_3 = \frac{(1-x_1)(1-x_2)}{(x_3-x_1)(x_3-x_2)(1-x_3)},$$

with the simulation points satisfying $x_1 < x_2 < x_3$. This is of the expected form (15).

Applying Theorem 2 we have $V_{\min} = (c_1 + c_2 + c_3)^2\sigma^2/N = S^2\sigma^2/N$, say. As all the c's are positive, we minimize S^2 if we minimize S. Now

$$\partial S/\partial x_1 \mid_{x_1=0} = 1 + 2(1-x_3)/[x_2^2(x_3 - x_2)] > 0$$

so the optimal value for x_1 is $x_1^* = 0$. We set x_1 to this value in S. We then get

$$\partial S(x_1^*, x_2, x_3)/\partial x_3 = \frac{x_2(x_2-x_3)^2 - 2(1-x_3)^2}{x_2(x_2-x_3)^2(1-x_3)^2}.$$

This equals zero if x_3 is equal to

$$x_3^*(x_2) = \frac{2 - x_2^2 - (2x_2)^{\frac{1}{2}}(1-x_2)}{2-x_2}.$$

Solving the equation

$$\partial S[x_1^*, x_2, x_3^*(x_2)]/\partial x_2 \equiv \frac{2^{\frac{1}{2}}(2^{\frac{1}{2}}x_2^{\frac{3}{2}} + 3x_2 - 1)}{x_2^{\frac{3}{2}}(1-x_2)^2} = 0$$

yields $x_2^* = 2 - \sqrt{3}$. We find therefore that the optimal input values are

$$x_1^* = 0 \quad x_2^* = 2 - \sqrt{3} \simeq 0.268 \quad x_3^* = 6 - 3\sqrt{3} \simeq 0.804$$

The optimal number of runs at these three simulation points follow from (16):

$$\begin{aligned} m_1^* &= (\tfrac{4}{9} - \tfrac{2\sqrt{3}}{9})N \simeq .0595N \\ m_2^* &= \tfrac{1}{6}N \simeq 0.1667N \\ m_3^* &= (\tfrac{2\sqrt{3}}{9} + \tfrac{7}{18})N \simeq 0.7738N. \end{aligned} \tag{22}$$

Thus the bulk of the simulation effort is at the highest traffic intensity setting.

The minimized variance follows from (17):

$$V_{\min} = (3\sqrt{3} + 6)^2\sigma^2/N \simeq 125.35\sigma^2/N.$$

An interesting aspect of this solution is that we do not make runs at traffic intensities close to unity (the highest setting is 0.804). This contrasts with standard polynomial situations where the end points of the range of interest are always used (see Kiefer and Wolfowitz, 1959). This becomes even more marked had we fitted the two-parameter model (first order metamodel)

$$z_i = (\gamma_0 + \gamma_1 x_i)(1 - x_i) + \delta_i \tag{23}$$

instead of (8). The coefficient of interest is then

$$p = \lim_{x \to 1}(\gamma_0 + \gamma_1 x) = \gamma_0 + \gamma_1, \tag{24}$$

so the variance of the estimator of this quantity, namely

$$Var(\hat{p}) = Var(\hat{\gamma}_0 + \hat{\gamma}_1), \tag{25}$$

is to be minimized. The optimal solution for this case sets the simulation points well away from $x = 1$:

$$x_1^* = 0 \quad and \quad x_2^* = 2 - \sqrt{2} \simeq 0.586.$$

The corresponding run allocations are

$$m_1^* = (\frac{1}{2} - \frac{\sqrt{2}}{4})N \simeq .1464N, \quad m_2^* = (\frac{1}{2} + \frac{\sqrt{2}}{4})N \simeq 0.8536N.$$

The minimized variance has value $(2\sqrt{2}+2)^2\sigma^2/N \simeq 23.31\sigma^2/N$.

Finally it is interesting to compare these settings with less guarded ones. Suppose in the three-parameter version (second order metamodel) we had used the same number of runs ($m_i = m$) at each of three equally spaced points: $x_1 = 0$, $x_2 = 0.45$, $x_3 = 0.9$. Then $Var(\hat{p}) = 555.74\sigma^2/N$, which is more than four times that of the minimum. The two-parameter case is even more dramatic. If we had used the same number of runs at each of the two points $x_1 = 0$ and $x_2 = 0.9$, then $Var(\hat{p}) \simeq 246.94\sigma^2/N$, which is more than ten times that of the minimum.

4.2 Simulation Results

To test the above theoretical analysis, we carried out two simulation experiments involving the M/M/1 queue. It is well known that the behaviour of even this elementary queue is hard to simulate, when the traffic load is near saturation. It is thus a good test model, especially as its analytical properties are well known, so that simulation results can be readily compared with theoretical results.

In the first simulation we fitted the three-coefficient model (8), whilst in the second simulation we fitted the two-coefficient model (23). The underlying true expected value that is being estimated is given by (2), so the correct coefficient values are $\gamma_0 = 0$, $\gamma_1 = 1$ and $\gamma_2 = 0$. Thus consistent estimators of the coefficients are obtained by fitting either metamodel, and the true value of p, defined in (9) is unity in both experiments. The results of the two experiments are summarized in Table 1. The confidence intervals calculated from the 100 replicates (described below) are given in the table. With the run lengths used, bias in the estimates is not a problem; in all cases the confidence intervals include the correct value of p.

For the three-coefficient model we first carried out a non-optimized version of the simulation. This was made up of five runs at each of three traffic intensities. The service rate was set at unity in all cases. The traffic intensities were therefore equal to the arrival rates in all cases, and were set at $x_1 = 0.001$, $x_2 = 0.450$, $x_3 = 0.900$, with very nearly equal numbers of customers, respectively: 6666, 6667 and 6667 simulated in each run of the given x value. The average waiting times of customers in each of the fifteen runs was recorded and the regression model (3) was then fitted to the fifteen observations, yielding the estimated value $\hat{p} = \hat{\gamma}_0 + \hat{\gamma}_1 + \hat{\gamma}_2$ as the response. This experiment was replicated 100 times, giving 100 independent $\hat{p}$ estimates. The sample mean and variance of these 100 values were 0.9812 and 0.01615 respectively.

An optimized version of this experiment was then carried out: again five runs at each of three traffic intensities were made, now using the optimum settings $x_1 = 0.001$, $x_2 = 0.268$, $x_3 = 0.804$ as found previously. The corresponding numbers of customers simulated were now respectively 1190, 3333, 15476 in each run of the given x value, using (22) with $N = 20,000$; also see the discussion of b_i and m_i immediately following (20). Again 100 replicates of the experiment ware made. The sample mean and variance of the 100 $\hat{p}$ values were 0.9986 and 0.00351 respectively.

Comparing the sample variances in the optimized and non-optimized versions, we see that the over four-fold reduction in the variability is very much that predicted by the theory (see the end of Section 5.1). Moreover this represents a genuine saving as the computing times were almost identical: 165.18 seconds and 165.70 seconds respectively for the non-optimized and the optimized experiments, running under compiled C-code on an Intel i860 processor.

Insert Table1 about here.

The second case was like the first, except that we fitted the two-coefficient model (23) and estimated (24), so that the variance to be minimized is (25). This time five runs were made at each of two settings in the non-optimized case, with 10,000 customers in each run at $x = 0.001$, and 10,000 customers also at $x = 0.900$. Again 100 replicates of the experiment were made. The sample mean and variance of the 100 $\hat{p}$ values were 0.9867 and 0.008363 respectively. In the optimized case 2,928 customers were simulated in each of five runs at $x = 0.001$, and 17,072 in each of five runs at $x = 0.586$. In this case the sample mean and variance of the 100 $\hat{p}$ values were 0.9918 and 0.0005334 respectively. Total computing times for the non-optimized and optimized cases were 163.73 and 163.92 seconds respectively. The gain in efficiency of over ten-fold is again much as predicted by the theory.

5 Conclusions and Future Research

We have shown the potential savings in using optimal design principles in regression metamodelling of queueing simulations. We have focused on the basic method of approach and for clarity we have deliberately avoided numerous side issues or extensions. However, there are some points to bear in mind, and several items that might be addressed in future work.

Terminating systems: Terminating queueing systems may not reach steady state. Then neither the variance nor the mean of waiting time explodes as the traffic load goes to unity. In particular the variance of the average waiting time per run does not follow the unbounded behaviour of $Var(\epsilon) \simeq (1-x)^{-4}\sigma^2$ as $x \to 1$. In this situation the optimal values of x may include both limits.

Finite run length: This point is closely connected with the previous one. Finite run length imposes a bound on the degree of extreme behaviour that can occur, even when the traffic load reaches the steady-state saturation level. Asmussen (1992) shows how different queueing behaviour will be observed, depending on how run length is increased as the traffic load is increased. In particular, Asmussen (1992, p.91) observes that excessively long runs are necessary to obtain steady-state behaviour as the traffic load approaches unity. See also Abate and Whitt (1987) and Whitt (1989). An advantage of our approach is that simulations are carried out away from the saturation level, so that such problems may not be so acute.

Unknown variances: This paper assumes that $Var(\epsilon)$ is a known function of σ^2; see (5). In practice this knowledge may not be available. Then a pilot study might be used to estimate how the variance depends on the input x value. This paper illustrates that substantial savings are possible even though approximate variance relations are used; in other words, our results are robust. If response variances are estimated, a nonlinear regression estimator of p results. Then jacknifing may be applied; see Kleijnen, Karremans, Oortwijn and Van Groenendaal (1987).

Validation of the regression metamodel: A lack of fit test and cross-validation may be used; see Kleijnen (1992). Validation is related to the call of Kiefer and Wolfowitz (1959, p.283) to 'choose the polynomial of correct degree'.

Other optimality criteria: We considered minimization of $Var(\hat{p})$, but the literature gives at least four more criteria, such as D-optimality ; see Kiefer and Wolfowitz (1959, pp. 285- 290) and Kleijnen (1987, pp. 335-336).

Sensitivity analysis: If we wish to know what happens to the response y when we change the input x, we are not then interested in the parameter γ_0. In this situation the linear function p in (9) needs to be changed by omitting

γ_0 from the expressions (10) and (24) used in the examples.

Non-linear regression analysis: Knowledge of analytical queueing models may suggest the use of non-linear regression metamodels. See the references in Kleijnen (1987, p.161). Other functions might be used to specify the dependence of the variance on the input.

Quantile estimation: In practice, managers may be more interested in (say) the 0.90 quantile rather than the mean; see Kleijnen (1987, p.426).

Inverse or control questions: Given y, what should be the value of x? Also see Kleijnen (1987, pp. 224-225).

Known heavy and light traffic results: Heavy traffic results are known for some queueing networks; the light traffic mean waiting time is simply the mean service time. Reiman et al. (1992) make the regression model pass through the known light and heavy traffic means (Constrained Least Squares). The light traffic end is usually easy to handle, whether one uses exact results or not. Heavy traffic results are certainly relevant and could have been incorporated in our analysis. But for practical systems they will often be unknown, so we have not assumed this knowledge in our analysis.

Multiple experimental factors: We assumed a single factor, namely traffic rate. Other factors such as the queueing priority rule may need to be taken into account; see Kleijnen (1987).

Variance reduction techniques: We have not considered use of common and antithetic seeds, control variates (see Kleijnen, 1975b, 1992), conditional sampling (see Cheng and Traylor 1993), and so on.

Estimating the exact saturation point: In complex systems this point may be unknown; for a case study see Griffiths and Williams (1989).

6 Acknowledgment

The second author was partly financed by the UK-Dutch Joint Scientific Research Programme of The British Council and the Netherlands Organization for Scientific Research (NWO).

Appendix A

Proof of Theorem 1: Consider first the determinant of **A**. This can be written as

$$\mid \mathbf{A} \mid = \sum_{(i_1, i_2, \ldots, i_{k+1}) \in P} \pm a_{1 i_1} a_{2 i_2} \ldots a_{k+1, i_{k+1}}$$

where P is the set of all permutations of the numbers $(1, 2, \ldots, k+1)$, and the sign of the term is positive or negative according to whether the permutation is even or odd. Now each element a_{ij} of A is of the form (12). Thus when fully expanded, each individual term will be of the form

$$\prod_{r=1}^{k+1} \{ n_{l_r} x_{l_r}^{r+i_r} \} \tag{26}$$

where each factor in braces arises from a given term in (12), this given term being indexed by l_r, and where the overall product is associated with the particular permutation $(i_1, i_2, \ldots, i_{k+1})$. Consider now any such product (26) and suppose $l_s = l_t = l$ say, for some given pair s, t where with no loss of generality we assume $s < t$. In other words, the variable n_l appears more than once in the product (26). We write the permutation associated with this product as $\pi = (i_1, \ldots, i_s, \ldots, i_t, \ldots, i_{k+1})$. We can pair this product with another product, which we call its *dual*, in a unique way. We simply take the permutation $\pi' = (i_1, ..., i_t, ..., i_s, ..., i_{k+1})$ obtained from π by interchanging i_s and i_t. The dual contains exactly the same sequence of factors in braces from each a_{r,i_r}. The only difference occurs in the two factors corresponding to i_s and i_t. These will be

$$g_s = \{n_l x_l^{s+i_s}\} \quad and \quad g_t = \{n_l x_l^{t+i_t}\}$$

in the product arising from π, and will be

$$h_s = \{n_l x_l^{s+i_t}\} \quad and \quad h_t = \{n_l x_l^{t+i_s}\}$$

in the dual arising from π'.

Now observe that the product of these two factors is the same: $g_s g_t = h_s h_t$. Thus the original product and the dual are equal in magnitude, as the remaining factors in each are unchanged. But only one interchange of symbols was made, in constructing π' from π. Thus one permutation is even, and one is odd. The signs of the two products are therefore opposite, and they cancel in their overall contribution to $\mid \mathbf{A} \mid$.

Moreover the dual of the dual is the original, so the set of all products separates into disjoint pairs. The above argument therefore applies separately to each pair. This means that the only non-zero terms that are not annihilated are those where each m_i appears once and once only. We have therefore proved that $\mid \mathbf{A} \mid$ has the form:

$$\mid \mathbf{A} \mid = d(x_1, ..., x_{k+1}) \prod_{i=1}^{k+1} n_i$$

where d is independent of the m's and r's. The cofactors of $\mathbf{A}$, $\mid \mathbf{A}_{ij} \mid$, share a similar property. An almost identical argument shows that no nonzero term in any cofactor can contain any n_i more than once. However, each cofactor is only of order k, so that precisely one n_i will be missing from each non-zero term. We have therefore that $|\mathbf{A}_{ij}|$ has the form

$$|\mathbf{A}_{ij}| = \sum_{s=1}^{k+1} d_{ijs}(x_1, ..., x_{k+1}) n_s^{-1} \prod_{t=1}^{k+1} n_t$$

where the d_{ijs} are expressions independent of the m's and r's.

The i, jth element of the inverse of $\mathbf{A}$ is therefore

$$(\mathbf{A}^{-1})_{ij} = |\mathbf{A}_{ji}| / |\mathbf{A}| = \sum_{s=1}^{k+1} \frac{d_{jis}(x_1, ..., x_{k+1})}{d(x_1, ..., x_{k+1})} n_s^{-1};$$

that is, $\mathbf{A}^{-1}$ is of the form (14).

Finally, the argument does not depend on the fact that $\sum_{i=1}^{n} m_i = N$. If we fix one m_i and let $m_j \to \infty$ for all $j \neq i$, then $\mathbf{A}^{-1} \to \mathbf{B}_i/[m_i r^2(x_i)]$. As $\mathbf{A}^{-1}$ is a covariance matrix, $\mathbf{B}_i$ is positive semidefinite.

This completes the proof of Theorem 1. □

References

Abate, J. and Whitt, W. (1994), "Transient behavior of the M/G/1 workload process." Operations Research, 42, 4, July-August, 750-764.

Aitken, A.C. (1964), Determinants and Matrices. Oliver and Boyd, Edinburgh.

Asmussen, S. (1992), "Queueing Simulation in Heavy Traffic." Math. of Operations Research, 17, 84-111.

Cheng, R.C.H. (1976), "A note on the effect of initial conditions on a simulation run." Operational Research Quarterly, 27, 2, ii, 467-470.

Cheng, R.C.H. (1990), "Fitting parametric models by conditional simulation." In Proceeding of the 1990 Winter Simulation Conference (eds. O. Balci, R.P. Sadowski and R.E. Nance), IEEE, Piscataway, New Jersey, 333-336.

Cheng, R.C.H. and Traylor, L. (1993), "Conditional simulation techniques for estimating the behaviour of nearly saturated queues.", UKSS Conference Proceedings 1993, 105-109.

Griffiths, J.D. and Williams, J.E. (1984), "Traffic studies on the Severn Bridge." Traffic Engineering and Control, 25(5), 268-271, 274.

Kiefer, J. and Wolfowitz, J. (1959), " Optimum designs in regression problems." Annals Mathematical Statistics, 30, 271-294.

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

Kleijnen, J.P.C. (1975a), "A comment on Blanning's metamodel for sensitivity analysis: the regression metamodel in simulation." Interfaces, 5, no. 3, May, 21-23.

Kleijnen, J.P.C. (1975b), "Antithetic variates, common random numbers and optimum computer time allocation." Management Science, Application Series, 21, no. 10, June, 1176-1185.

Kleijnen J.P.C. (1974/1975), Statistical techniques in simulation (part I/part II). Marcel Dekker, New York.

Kleijnen, J.P.C., Karremans P.C.A., Oortwijn, W.K. and Van Groenendaal W.J.H. (1987), "Jackknifing estimated weighted least squares: JEWLS." Communications in Statistics, Theory and Methods, 16, no. 3, 747-764.

Kleijnen, J.P.C. (1992), "Regression metamodels for simulation with common random numbers: comparison of validation tests and confidence intervals." Management Science, 38, no. 8, August, 1164-1185.

Reiman M.I., Simon B., and Willie J.S. (1992), "Simterpolation: a simulation based interpolation approximation for queueing systems." Operations Research, 40, no. 4, July-August, 706-723

Vollebregt, T.A.J. (1994a), "Experimental design for simulation", Working paper, Dept of Management, Univ. of Canterbury, New Zealand.

Vollebregt, T.A.J. (1994b), "Combining optimal design and sequential analysis", Working paper, Dept of Management, Univ. of Canterbury, New Zealand.

Whitt, W. (1989), "Planning queueing simulations." Management Science, 35, no. 11, November, 1341-1366.

Table1

Estimating E(W) in the M/M/1Queue

1a. *Using* $p = \gamma_0 + \gamma_1 + \gamma_2$. *Correct value is* $p = 1$

Non − Optimized		*Optimized*	
x	# *of customers/per run*	x	# *of customers/per run*
0.001	6,666	0.001	1,190
0.450	6,667	0.268	3,333
0.900	6,667	0.804	15,476

\# *of runs at each* x : 5

Results from 100 *simulations* :

Mean $\hat{p}$ 0.9812	*Mean* $\hat{p}$ 0.9986
Var $\hat{p}$ 0.01615	*Var* $\hat{p}$ 0.003510

95% *Confidence Interval for* p :

(.9563, 1.0061)	(.9870, 1.0102)

Time for the 100 *simulations* :

165.18 *seconds*	165.70 *seconds*

1b. *Using* $p = \gamma_0 + \gamma_1$. *Correct value is* $p = 1$

Non − Optimized		*Optimized*	
x	# *of customers/per run*	x	# *of customers/per run*
0.001	10,000	0.001	2,928
0.900	10,000	0.586	17,072

\# *of runs at each* x : 5

Results from 100 *simulations* :

Mean $\hat{p}$ 0.9867	*Mean* $\hat{p}$ 0.9918
Var $\hat{p}$ 0.008363	*Var* $\hat{p}$ 0.0005334

95% *Confidence Interval for* p :

(.9688, 1.0046)	(.9918, 1.0008)

Time for the 100 *simulations* :

163.73 *seconds*	163.92 *seconds*

No.	Author(s)	Title
9493	S. Eijffinger, M. van Rooij, and E. Schaling	Central Bank Independence: A Paneldata Approach
9494	S. Eijffinger and M. van Keulen	Central Bank Independence in Another Eleven Countries
9495	H. Huizinga	The Incidence of Interest Withholding Taxes: Evidence from the LDC Loan Market
9496	V. Feltkamp, S. Tijs and S. Muto	Minimum Cost Spanning Extension Problems: The Proportional Rule and the Decentralized Rule
9497	J.P.J.F. Scheepens	Financial Intermediation, Bank Failure and Official Assistance
9498	A.L. Bovenberg and R.A. de Mooij	Environmental Tax Reform and Endogenous Growth
9499	J. Ashayeri, R. Heuts, A. Jansen and B. Szczerba	Inventory Management of Repairable Service Parts for Personal Computers: A Case Study
94100	A. Cukierman and S. Webb	Political Influence on the Central Bank - International Evidence
94101	G.J. Almekinders and S.C.W. Eijffinger	The Ineffectiveness of Central Bank Intervention
94102	R. Aalbers	Extinction of the Human Race: Doom-Mongering or Reality?
94103	H. Bester and W. Güth	Is Altruism Evolutionarily Stable?
94104	H. Huizinga	Migration and Income Transfers in the Presence of Labor Quality Externalities
94105	F.C. Drost, T.E. Nijman, and B.J.M. Werker	Estimation and Testing in Models Containing both Jumps and Conditional Heteroskedasticity
94106	V. Feltkamp, S. Tijs, and S. Muto	On the Irreducible Core and the Equal Remaining Obligations Rule of Minimum Cost Spanning Extension Problems
94107	D. Diamantaras, R.P. Gilles and P.H.M. Ruys	Efficiency and Separability in Economies with a Trade Center
94108	R. Ray	The Reform and Design of Commodity Taxes in the Presence of Tax Evasion with Illustrative Evidence from India
94109	F.H. Page	Optimal Auction Design with Risk Aversion and Correlated Information

No.	Author(s)	Title
94110	F. de Roon and C. Veld	An Empirical Investigation of the Factors that Determine the Pricing of Dutch Index Warrants
94111	P.J.-J. Herings	A Globally and Universally Stable Quantity Adjustment Process for an Exchange Economy with Price Rigidities
94112	V. Bhaskar	Noisy Communication and the Fast Evolution of Cooperation
94113	R.C. Douven and J.E.J. Plasmans	S.L.I.M. - A Small Linear Interdependent Model of Eight EU-Member States, the USA and Japan
94114	B. Bettonvil and J.P.C. Kleijnen	Identifying the Important Factors in Simulation Models with Many Factors
94115	H. Uhlig and N. Yanagawa	Increasing the Capital Income Tax Leads to Faster Growth
9501	B. van Aarle, A.L. Bovenberg and M. Raith	Monetary and Fiscal Policy Interaction and Government Debt Stabilization
9502	B. van Aarle and N. Budina	Currency Substitution in Eastern Europe
9503	Z. Yang	A Constructive Proof of a Unimodular Transformation Theorem for Simplices
9504	J.P.C. Kleijnen	Sensitivity Analysis and Optimization of System Dynamics Models: Regression Analysis and Statistical Design of Experiments
9505	S. Eijffinger and E. Schaling	The Ultimate Determinants of Central Bank Independence
9506	J. Ashayeri, A. Teelen and W. Selen	A Production and Maintenance Planning Model for the Process Industry
9507	J. Ashayeri, A. Teelen and W. Selen	Computer Integrated Manufacturing in the Chemical Industry: Theory & Practice
9508	A. Mountford	Can a Brain Drain be Good for Growth?
9509	F. de Roon and C. Veld	Announcement Effects of Convertible Bond Loans Versus Warrant-Bond Loans: An Empirical Analysis for the Dutch Market
9510	P.H. Franses and M. McAleer	Testing Nested and Non-Nested Periodically Integrated Autoregressive Models
9511	R.M.W.J. Beetsma	The Political Economy of a Changing Population
9512	V. Kriman and R.Y. Rubinstein	Polynomial Time Algorithms for Estimation of Rare Events in Queueing Models

No.	Author(s)	Title
9513	J.P.C. Kleijnen, and R.Y. Rubinstein	Optimization and Sensitivity Analysis of Computer Simulation Models by the Score Function Method
9514	R.D. van der Mei	Polling Systems with Markovian Server Routing
9515	M. Das	Extensions of the Ordered Response Model Applied to Consumer Valuation of New Products
9516	P.W.J. De Bijl	Entry Deterrence and Signaling in Markets for Search Goods
9517	G. Koop, J. Osiewalski and M.F.J. Steel	The Components of Output Growth: A Cross-Country Analysis
9518	J. Suijs, H. Hamers and S. Tijs	On Consistency of Reward Allocation Rules in Sequencing Situations
9519	R.F. Hartl and P.M. Kort	Optimal Input Substitution of a Firm Facing an Environmental Constraint
9520	A. Lejour	Cooperative and Competitive Policies in the EU: The European Siamese Twin?
9521	H.A. Keuzenkamp	The Econometrics of the Holy Grail: A Critique
9522	E. van der Heijden	Opinions concerning Pension Systems. An Analysis of Dutch Survey Data
9523	P. Bossaerts and P. Hillion	Local Parametric Analysis of Hedging in Discrete Time
9524	S. Hochgürtel, R. Alessie and A. van Soest	Household Portfolio Allocation in the Netherlands: Saving Accounts versus Stocks and Bonds
9525	C. Fernandez, J. Osiewalski and M.F.J. Steel	Inference Robustness in Multivariate Models with a Scale Parameter
9526	G.-J. Otten, P. Borm, T. Storcken and S. Tijs	Decomposable Effectivity Functions
9527	M. Lettau and H. Uhlig	Rule of Thumb and Dynamic Programming
9528	F. van Megen, P. Borm, and S. Tijs	A Perfectness Concept for Multicriteria Games
9529	H. Hamers	On the Concavity of Delivery Games
9530	V. Bhaskar	On the Generic Instability of Mixed Strategies in Asymmetric Contests
9531	E. Canton	Efficiency Wages and the Business Cycle

No.	Author(s)	Title
9532	J.J.G. Lemmen and S.C.W. Eijffinger	Financial Integration in Europe: Evidence from Euler Equation Tests
9533	P.W.J. De Bijl	Strategic Delegation of Responsibility in Competing Firms
9534	F. de Jong and T. Nijman	High Frequency Analysis of Lead-Lag Relationships Between Financial Markets
9535	B. Dutta, A. van den Nouweland and S. Tijs	Link Formation in Cooperative Situations
9536	B. Bensaid and O. Jeanne	The Instability of Fixed Exchange Rate Systems when Raising the Nominal Interest Rate is Costly
9537	E.C.M. van der Heijden, J.H.M. Nelissen and H.A.A. Verbon	Altruism and Fairness in a Public Pension System
9538	L. Meijdam and H.A.A. Verbon	Aging and Public Pensions in an Overlapping-Generations Model
9539	H. Huizinga	International Trade and Migration in the Presence of Sector-Specific Labor Quality Pricing Distortions
9540	J. Miller	A Comment on Holmlund & Lindén's "Job Matching, Temporary Public Employment, and Unemployment"
9541	H. Huizinga	Taxation and the Transfer of Technology by Multinational Firms
9542	J.P.C. Kleijnen	Statistical Validation of Simulation Models: A Case Study
9543	H.L.F. de Groot and A.B.T.M. van Schaik	Relative Convergence in a Dual Economy with Tradeable and Non-Tradeable Goods
9544	C. Dustmann and A. van Soest	Generalized Switching Regression Analysis of Private and Public Sector Wage Structures in Germany
9545	C. Kilby	Supervision and Performance: The Case of World Bank Projects
9546	G.W.J. Hendrikse and C.P. Veerman	Marketing Cooperatives and Financial Structure
9547	R.M.W.J. Beetsma and A.L. Bovenberg	Designing Fiscal and Monetary Institutions in a Second-Best World
9548	R. Strausz	Collusion and Renegotiation in a Principal-Supervisor-Agent Relationship
9549	F. Verboven	Localized Competition, Multimarket Operation and Collusive Behavior

No.	Author(s)	Title
9550	R.C. Douven and J.C. Engwerda	Properties of *N*-person Axiomatic Bargaining Solutions if the Pareto Frontier is Twice Differentiable and Strictly Concave
9551	J.C. Engwerda and A.J.T.M. Weeren	The Open-Loop Nash Equilibrium in LQ-Games Revisited
9552	M. Das and A. van Soest	Expected and Realized Income Changes: Evidence from the Dutch Socio-Economic Panel
9553	J. Suijs	On Incentive Compatibility and Budget Balancedness in Public Decision Making
9554	M. Lettau and H. Uhlig	Can Habit Formation be Reconciled with Business Cycle Facts?
9555	F.H. Page and M.H. Wooders	The Partnered Core of an Economy
9556	J. Stennek	Competition Reduces X-Inefficiency. A Note on a Limited Liability Mechanism
9557	K. Aardal and S. van Hoesel	Polyhedral Techniques in Combinatorial Optimization
9558	R.M.W.J. Beetsma and A.L. Bovenberg	Designing Fiscal and Monetary Institutions for a European Monetary Union
9559	R.M.W.J. Beetsma and A.L. Bovenberg	Monetary Union without Fiscal Coordination May Discipline Policymakers
9560	R. Strausz	Delegation of Monitoring in a Principal-Agent Relationship
9561	A. Lejour	Social Insurance and the Completion of the Internal Market
9562	J. Bouckaert	Monopolistic Competition with a Mail Order Business
9563	H. Haller	Household Decisions and Equilibrium Efficiency
9564	T. Chou and H. Haller	The Division of Profit in Sequential Innovation Reconsidered
9565	A. Blume	Learning, Experimentation, and Long-Run Behavior in Games
9566	H. Uhlig	Transition and Financial Collapse
9567	R.C.H. Cheng and J.P.C. Kleijnen	Optimal Design of Simulation Experiments with Nearly Saturated Queues