Statistica Sinica 2(1992), 413-437

EFFICIENT SEMI-LATIN SQUARES

R. A. Bailey

University of London, Goldsmiths' College

Abstract: Consider an $n \times n$ square array in which each small square is divided into k plots. *A semi-Latin square* is an allocation of nk treatments to the plots of such an array so that each treatment occurs once in each row and once in each column. Several different practical situations are discussed which all lead to this same abstract structure.

There are two reasonable models for data from semi-Latin squares. Under the first, all semi-Latin squares are equally efficient, while under the second there is a wider range of efficiencies. Attention is focused on the problem of finding efficient semi-Latin squares for the second model.

There is a family of semi-Latin squares called *Trojan* squares, which are known to be optimal, as are certain squares derived from the Trojan squares. Unfortunately, these do not exist for all pairs of values of n and k. Recent agricultural experiments have required efficient semi-Latin squares for some of these other values of n and k. New designs for these values are presented and their efficiencies and possible optimality discussed.

Key words and phrases: Efficiency factor, incomplete block design, Latin square, optimal design, semi-Latin square, Trojan square.

1. Combinatorial Aspects of Semi-Latin Squares

There is some confusion between the combinatorial object known as a semi-Latin square and various uses to which such an object can be put, both in designed experiments and elsewhere. In this section we are concerned only with the combinatorics.

Definition. Let Ω be a set of n^2k points which is divided into n rows and n columns in such a way that the intersection of each row with each column contains k points. Suppose that nk symbols are allocated to the points, n points to each symbol. If each symbol occurs once in each row and once in each column then Ω is a $(n \times n)/k$ *semi-Latin squre.*

Examples of semi-Latin squares are in Figures 1(a), 3(a), 4(a), 5, 7 and 12.

One simple way to construct an $(n \times n)/k$ semi-Latin square is to take an $n \times n$ Latin square and replace each letter by k new symbols. This gives an *inflated Latin square*. An example is in Figure 1(a).

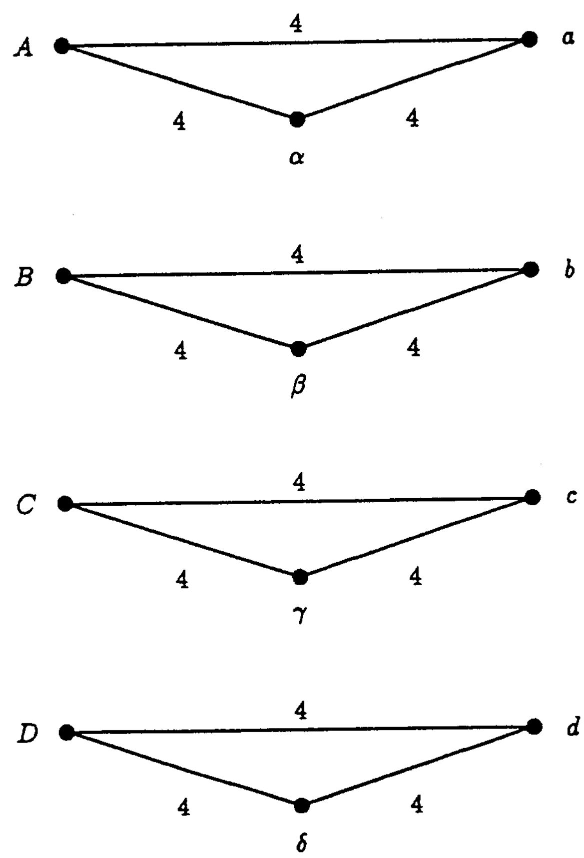
Let R, C and S be the partitions of Ω into rows, columns and symbols respectively. In the notation of Bailey (1985, 1989), the *infimum* $R \wedge C$ of R and C is a partition into n^2 classes of size k—the intersections of rows with columns. It is convenient to call these classes *blocks*. Similarly, $R \wedge S = C \wedge S = E$, where E denotes the trivial partition of Ω into n^2k singletons. The other trivial partition U of Ω contains a single class, the whole of Ω. We have

$$R \vee C = R \vee S = C \vee S = U,$$

Figure 1. Inflated Latin square with $n = 4$ and $k = 3$

$A\ \alpha\ a$	$B\ \beta\ b$	$C\ \gamma\ c$	$D\ \delta\ d$
$D\ \delta\ d$	$A\ \alpha\ a$	$B\ \beta\ b$	$C\ \gamma\ c$
$C\ \gamma\ c$	$D\ \delta\ d$	$A\ \alpha\ a$	$B\ \beta\ b$
$B\ \beta\ b$	$C\ \gamma\ c$	$D\ \delta\ d$	$A\ \alpha\ a$

(a) The semi-Latin square



(b) Its variety-concurrence graph

where $\vee$ denotes supremum. (The supremum of two partitions has classes which consist of whole classes of each partition, and which are as small as possible subject to this condition.)

Tjur (1984) defines two partitions, or factors, F_1 and F_2, to be *geometrically orthogonal* if, within each class g of their supremum $F_1 \vee F_2$, every F_1-class f_1 meets every F_2-class f_2 and the size of their intersection $f_1 \cap f_2$ is equal to $|f_1||f_2|/|g|$. Statisticians usually abbreviate 'geometrically orthogonal' to 'orthogonal'. Thus R is orthogonal to C, R is orthogonal to S, and C is orthogonal to S. However, the partitions $R \wedge C$ and S may or may not be orthogonal to each other.

Definition. A semi-Latin square is *orthogonal* if its partitions $R \wedge C$ and S are mutually orthogonal.

Theorem 1.1. *A semi-Latin square is orthogonal if and only if it is an inflated Latin square.*

Proof. Let $L = (R \wedge C) \vee S$. Suppose that the semi-Latin square Ω is orthogonal. Let ℓ be a class of L. Then ℓ consists of a union of whole blocks (each of size k), and a union of whole symbols (each of size n). Within ℓ, every block meets every symbol in a single point, because no symbol occurs more than once in a block. Thus ℓ consists of n whole blocks, which between them contain all the nk points with k given symbols, and which therefore lie in different rows and columns. Hence the classes of L form an $n \times n$ Latin square (on the n^2 blocks), of which Ω is an inflation.

Conversely, if Ω is obtained by inflating an $n \times n$ Latin square Λ then the classes of L are the letters of Λ and the orthogonality is obvious.

Thus, the partitions U, R, C, $R \wedge C$, L, S and E of an orthogonal semi-Latin square form an *orthogonal block structure*, as defined by Bailey (1984, 1985, 1989, 1991), with Hasse diagram as shown in Figure 2.

Suppose that there is a partition $\Omega_1 \cup \cdots \cup \Omega_k$ of Ω and a partition $S_1 \cup \cdots \cup S_k$ of the symbols such that, for $i = 1, \ldots, k$,

- the set S_i contains n symbols;
- the set Ω_i contains n^2 points, all of whose symbols are in S_i;
- the restrictions R_i and C_i of R and C to Ω_i make $(\Omega_i, R_i, C_i, S_i)$ into a Latin square Λ_i.

Then Ω is called the *superposition* of the Latin squares $\Lambda_1, \ldots, \Lambda_k$. If the squares $\Lambda_1, \ldots, \Lambda_k$ are identical except for the naming of the symbols then Ω is orthogonal. At the other extreme, if the squares $\Lambda_1, \ldots, \Lambda_k$ are mutually orthogonal then Ω is called a *Trojan square* by Darby and Gilbert (1958). Examples of Trojan squares are in Figures 3(a), 5 and 7.

Figure 2. Hasse diagram of an orthogonal $(n \times n)/k$ semi-Latin square

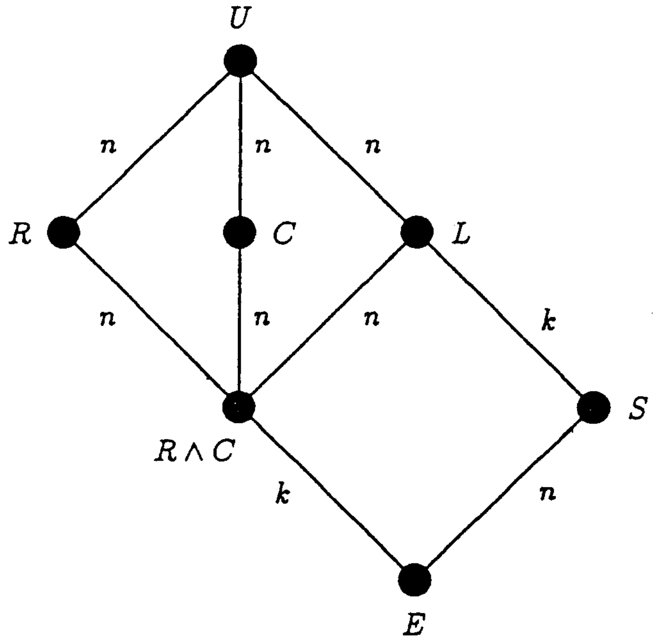
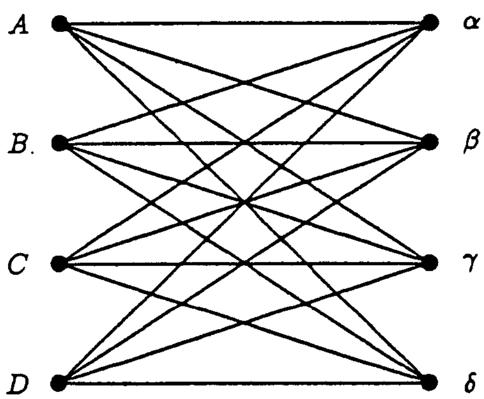


Figure 3. Trojan square with $n = 4$ and $k = 2$

$A\ \alpha$	$B\ \beta$	$C\ \gamma$	$D\ \delta$
$B\ \gamma$	$A\ \delta$	$D\ \alpha$	$C\ \beta$
$C\ \delta$	$D\ \gamma$	$A\ \beta$	$B\ \alpha$
$D\ \beta$	$C\ \alpha$	$B\ \delta$	$A\ \gamma$

(a) The semi-Latin square



(b) Its variety-concurrence graph

The operations of inflation and superposition can also be applied to semi-Latin squares. In the latter case the constituent semi-Latin squares do not need to have the same value of k.

If the rows and columns of Ω are ignored, the blocks and symbols form a binary incomplete-block design $\Delta(\Omega)$. The variety-concurrence graph $\mathcal{G}(\Omega)$ of this design is described by Patterson and Williams (1976a) and Paterson (1983). It has the symbols as vertices; the number of edges joining symbols s_1 and s_2 is the number of blocks in which s_1 and s_2 both occur. Thus the variety-concurrence graph of an orthogonal semi-Latin square consists of n complete graphs on k vertices, with all edges of multiplicity n; while that of a Trojan square is a complete k-partite graph on k sets of n vertices. Examples of these graphs are in Figures 1(b) and 3(b).

Theorem 1.2. *The variety-concurrence graph of an $(n\times n)/k$ semi-Latin square Ω is k-partite if and only if Ω is the superposition of k Latin squares.*

Proof. If Ω is the superposition of Latin squares $\Lambda_1, \ldots, \Lambda_k$ then let S_i be the set of symbols in Λ_i, for $i = 1, \ldots, k$. In $\mathcal{G}(\Omega)$ there are no edges between vertices in S_i, for any i. Hence $\mathcal{G}(\Omega)$ is k-partite.

Conversely, if $\mathcal{G}(\Omega)$ is k-partite then let $S_1, \ldots, S_k$ be the sets of vertices such that all edges of $\mathcal{G}(\Omega)$ join two of these sets. Since no two vertices in S_i are joined, the symbols in S_i must be in different blocks in any one row, and hence $|S_i| \leq n$. But $\sum_i |S_i| = nk$ and so $|S_i| = n$ for $i = 1, \ldots, k$. Thus the points of Ω whose symbols are in S_i form an $n \times n$ Latin square Λ_i, and so Ω is the superposition of $\Lambda_1, \ldots, \Lambda_k$.

Corollary 1.3. *An $(n \times n)/k$ semi-Latin square is Trojan if and only if its variety-concurrence graph is a complete k-partite graph with no multiple edges.*

Figure 4 shows a $(5 \times 5)/2$ semi-Latin square which is not a superposition; its graph is not bipartite.

2. Practical Use of Semi-Latin Squares

Semi-Latin squares arise in various practical applications. The first five examples are from real designed experiments.

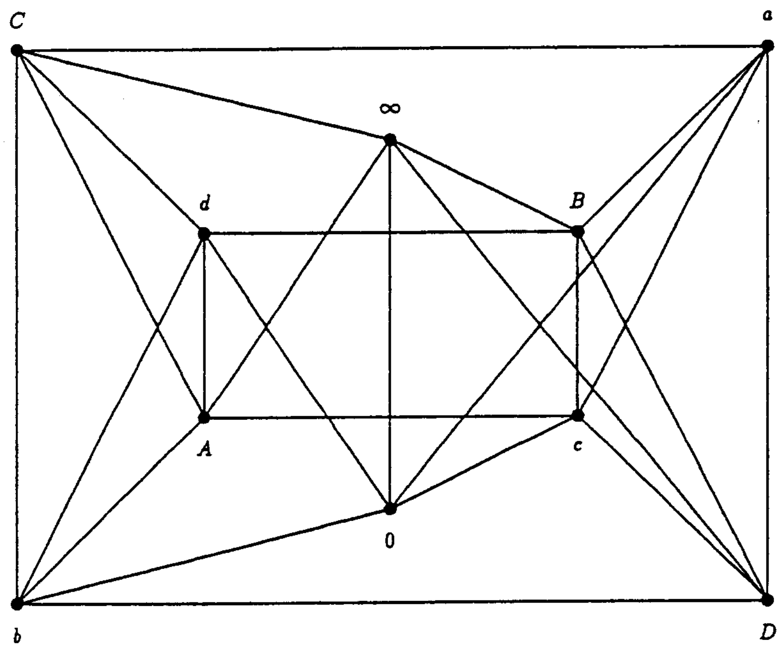
Example 1 (*Consumer testing*). A consumer research organization wishes to compare eight new brands of vacuum cleaner. The organization has bought one sample of each brand. Some housewives have agreed to compare the vacuum cleaners. Each housewife will use two vacuum cleaners in her home for a week and give each one a score. Thus at most four housewives can test cleaners in any one week. Moreover, to allow for housewife effects, it is best that each housewife test every cleaner, and therefore take part in the trial for four weeks. A design such as the one in Figure 3(a) is possible, with rows representing weeks, columns housewives and symbols cleaners.

This situation arises frequently in consumer testing, when only one object of each of nk brands is available. Only one of each brand can therefore be used at any one time, and the trial is completed most quickly if every brand is used in every time-period. Consumers test k objects per time-period in their own homes: typically $k = 2$ or $k = 3$. To eliminate consumer effects, each of n consumers participates in the trial for n weeks, with time-periods, consumers and brands forming the rows, columns and symbols of a semi-Latin square. If a further n time-periods are available for the trial, then a second semi-Latin square is used, either with the same n consumers or with n completely different consumers.

Figure 4. A semi-Latin square for $n = 5$ and $k = 2$ which is not a superposition

0	∞	A	d	D	a	C	b	B	c
B	d	0	b	A	C	a	c	D	∞
C	a	D	c	B	∞	0	d	A	b
A	c	C	∞	b	d	B	D	0	a
D	b	B	a	0	c	A	∞	C	d

(a) The semi-Latin square



(b) Its variety-concurrence graph

Example 2 (*Glasshouse crops*). In trials on glasshouse crops there are usually pronounced row and column effects. The glasshouse is typically rectangular, say

nk plots by n plots, with its long axis East-West. Because North-South variation is usually greater per distance than East-West variation, it is reasonable to divide the glasshouse into n rows and n columns, where each row is an East-West line of plots and each column consists of k contiguous North-South lines of plots. If there are nk treatments—such as varieties of lettuce, or sowing dates for celery, or factorial treatments on tomatoes—they can be applied according to the symbols of a semi-Latin square. The Trojan square in Figure 5 is suitable for 15 varieties of lettuce in a glasshouse which is 15 plots East-West by 5 plots North-South. Indeed, Trojan squares were originally developed for this application by Darby and Gilbert (1958).

Figure 5. Trojan square with $n = 5$ and $k = 3$

$A\ \alpha\ a$	$B\ \beta\ b$	$C\ \gamma\ c$	$D\ \delta\ d$	$E\ \varepsilon\ e$
$E\ \delta\ c$	$A\ \varepsilon\ d$	$B\ \alpha\ e$	$C\ \beta\ a$	$D\ \gamma\ b$
$D\ \beta\ e$	$E\ \gamma\ a$	$A\ \delta\ b$	$B\ \varepsilon\ c$	$C\ \alpha\ d$
$C\ \varepsilon\ b$	$D\ \alpha\ c$	$E\ \beta\ d$	$A\ \gamma\ e$	$B\ \delta\ a$
$B\ \gamma\ d$	$C\ \delta\ e$	$D\ \varepsilon\ a$	$E\ \alpha\ b$	$A\ \beta\ c$

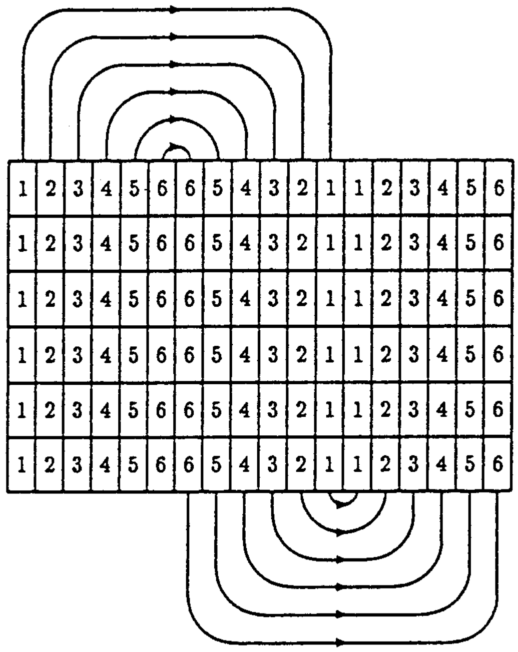
As reported by Rojas and White (1957), semi-Latin squares have been used for agricultural field trials in Mexico in much the same manner as Example 2. In this context they are sometimes called *modified* Latin squares. Rasch and Herrendörfer (1986) call them *pseudo*-Latin squares.

Example 3 (*Residual effects*). The effects of some treatments may persist during the next experiment. This is particularly true if the experimental units are trees, but it can also occur on arable crops if the treatments affect the soil directly, for example by inhibiting or encouraging nematode growth. Suppose that last year five varieties of potato were compared in five replicates. This year, a single standard variety is grown and ten chemicals are tested for their ability to control nematodes. Last year's varieties will affect the number of nematodes in the soil, but it is assumed that there is no interaction between those varieties and this year's chemicals.

Chemicals can be applied to smaller areas of land than varieties, so each plot from last year is split into two for the chemicals. Use of a semi-Latin square, with rows representing old replicates, columns varieties and symbols chemicals, ensures that each chemical occurs once in each old replicate and once on soil that had each variety last year. The semi-Latin square in Figure 4 could be used, but may not be the best choice.

Example 4 (*Sugar beet trials*). In sugar beet trials each plot is sometimes a single line of plants. A typical size is 0.5 metres wide by 3 metres long. The plots are arranged in an $n \times nk$ rectangle, with plots in the same row adjacent along long edges to keep the trial area compact. Thus plots in the same row are likely to have similar soil, and so it is sensible to use rows for blocking, especially if there are nk treatments. However, the drilling machine drills n plots in a row at once, travelling along the plant lines across the rows, then turning through 180° and returning in the reverse direction. Figure 6 shows the lines of drilling for $n = 6$ and $k = 3$. Plots with the same number are drilled by the same drill. Thus the n drills constitute a second blocking system, which should be taken into account in the design and analysis of the trial. There are precisely k plots in common to each row and drill, and so the structure in Figure 6 is abstractly identical to the row-and-column part of a $(n \times n)/k$ semi-Latin square. Applying the treatments according to the symbols of a semi-Latin square ensures that treatments are orthogonal to both rows and drills.

Figure 6. Drilling of sugar beet trials ($n = 6$ and $k = 3$) (not to scale — the area is 9m wide by 18m long)



Example 5 (*Food industry*). An interesting experiment to compare colour intensities of apple sauce was described by Harshbarger and Davis (1952). The treatments consisted of all combinations of 12 blends of apple sauce with four concentrations of cinnamon. Treatments could be stored for four different lengths of time. A $(4 \times 4)/3$ Trojan square was used in which rows, columns and symbols represented cinnamon concentrations, storage times and blends respectively. This ensured that each of the 48 treatments occurred once, and that both treatment

factors were orthogonal to storage times. Part of the interaction between blends and concentrations was totally confounded with storage times.

The *quotient* block design of a semi-Latin square is the incomplete-block design derived from it by ignoring the rows and columns: see Bailey (1988). Because the quotient block design of a Trojan square with $k = n-1$ is necessarily a rectangular lattice design, Harshbarger and Davis (1952) called their Trojan square a *latinized rectangular lattice.* Generalizing this idea, Williams (1986) called semi-Latin squares *latinized incomplete-block designs.*

Example 6 (*Fractional factorials*). Suppose that R, C and S are treatment factors with n, n and nk levels respectively. If there are no interactions between the factors, then the three main effects may be orthogonally estimated from the fractional replicate consisting of the n^2k combinations in a semi-Latin square.

Thus a semi-Latin square is an asymmetrical *orthogonal array* with n^2k assemblies (or plots), three constraints (factors R, C and S) and strength 2. This means that, for every pair of treatment factors, each pair of levels occurs equally often in the fraction.

In the language of design of experiments, a *factor* takes only one level on any one plot. A nature generalization is a *multi-factor*, which takes m levels on each plot, for some m. This idea is useful in diallel structures (with $m = 2$), in rectangular lattices (see Thompson (1984) and Bailey and Speed (1986)), and in experiments with neighbour effects from more than one side. Brickell (1984) defined an *orthogonal multi-array* by replacing constraints by multi-factors (with various values of m) in an orthogonal array with strength 2 and minimal number of plots. Thus an orthogonal multi-array $\mathrm{OMA}(p, n; m_1, m_2, \ldots, m_p)$ consists of n^2 plots and p multi-factors with $m_1, \ldots, m_p$ levels per plot, such that each pair of levels from each pair of multi-factors occurs on just one plot. In this language, an $(n \times n)/k$ semi-Latin square is an $\mathrm{OMA}(3, n;\ 1,\ 1,\ k)$.

Example 7 (*Message authentication*). Semi-Latin squares are also used as doubly perfect authentication schemes in communications theory: see Anthony et al. (1990). There are three source messages (R, C and S) and n^2 encoding rules, corresponding to the n^2 blocks. If the block in row i and column j is the current encoding rule, then source message R is transmitted as R_i, source message C is transmitted as C_j, and source message S is transmitted as a random symbol from this block. An interfering person, who does not know the current encoding rule and who wants to either transmit a false message or intercept a message and replace it with a false one, has chance only $1/n$ of being believed by the receiver to be the authentic transmitter. This chance is minimum over all message authentication schemes of this size.

3. Statistical Models and Analysis

In Examples 1–4 the symbols of the semi-Latin square are used for treatments of the current experiment, while rows and columns represent pre-existing conditions. I shall assume an additive model for the response y, with fixed treatment effects. Row and column effects should also be included. As S is orthogonal to both R and C, and $R \vee S = C \vee S = U$, it makes no difference to the analysis whether R and C are regarded as fixed or random.

Some people argue that the weeks in Example 1 should not be included in the model, because they are unlikely to differ among themselves. Moreover, they are incorporated in the design only because of the limited availability of the vacuum cleaners, not because they are believed to be an essential blocking system. In my view, the changing of the treatments at the end of each week induces differences between the weeks: for example, consumers may take stock of what they are doing and change their habits slightly. To see this more clearly, consider the situation where treatments can be applied only to whole weeks. It would be wrong to use within-week or within-week-and-consumer variability to assess treatment differences in that situation. In both situations it is the constraints on treatment application which force us to consider weeks as meaningful blocks, and hence to include them in the model.

A more controversial question is whether the factor $R \wedge C$ should be included in the model. A 'random effects' model of the type studied by Nelder (1965) and Speed and Bailey (1987) assumes that the expectation of y is a function of the allocated treatments and that the covariance matrix $\mathcal{C}$ of y is given by

$$\begin{aligned} \mathcal{C} &= \gamma_E I + \gamma_{R\wedge C}(J_{R\wedge C} - I) + \gamma_R(J_R - J_{R\wedge C}) \\ &\quad + \gamma_C(J_C - J_{R\wedge C}) + \gamma_U(J - J_R - J_C + J_{R\wedge C}). \end{aligned} \tag{1}$$

Here, I is the identity matrix, J the all-1s matrix, and the (ω, ω')-entry of J_R, J_C, and $J_{R\wedge C}$ is equal to 1 if plots ω and ω' are in the same row, column, block respectively, and to 0 otherwise. Equation (1) may be regarded as a *patterns of covariance* model, because the γ terms are covariances, which depend on the relationships between the two plots involved. This equation may fruitfully be rewritten in two other forms (see Speed (1987) and Speed and Bailey (1987)): the *components-of-variance* model

$$\mathcal{C} = \sigma_E^2 I + \sigma_{R\wedge C}^2 J_{R\wedge C} + \sigma_R^2 J_R + \sigma_C^2 J_C + \sigma_U^2 J, \tag{2}$$

where

$$\begin{array}{lcllllllll}
\sigma_E^2 & = & \gamma_E & - & \gamma_{R\wedge C} & & & & & \\
\sigma_{R\wedge C}^2 & = & & & \gamma_{R\wedge C} & - & \gamma_R & - & \gamma_C & + & \gamma_U \\
\sigma_R^2 & = & & & & & \gamma_R & & & - & \gamma_U \\
\sigma_C^2 & = & & & & & & & \gamma_C & - & \gamma_U \\
\sigma_U^2 & = & & & & & & & & & \gamma_U;
\end{array}$$

and the *spectral* form

$$\begin{aligned}
C &= \xi_E\left(I - k^{-1}J_{R\wedge C}\right) \\
&\quad + \xi_{R\wedge C}\left(k^{-1}J_{R\wedge C} - (nk)^{-1}J_R - (nk)^{-1}J_C + (n^2k)^{-1}J\right) \\
&\quad + \xi_R\left((nk)^{-1}J_R - (n^2k)^{-1}J\right) + \xi_C\left((nk)^{-1}J_C - (n^2k)^{-1}J\right) \\
&\quad + \xi_U(n^2k)^{-1}J, \qquad (3)
\end{aligned}$$

where

$$\begin{array}{lcllllllll}
\xi_E & = & \sigma_E^2 & & & & & & & \\
\xi_{R\wedge C} & = & \sigma_E^2 & + & k\sigma_{R\wedge C}^2 & & & & & \\
\xi_R & = & \sigma_E^2 & + & k\sigma_{R\wedge C}^2 & + & nk\sigma_R^2 & & & \\
\xi_C & = & \sigma_E^2 & + & k\sigma_{R\wedge C}^2 & & & + & nk\sigma_C^2 & \\
\xi_U & = & \sigma_E^2 & + & k\sigma_{R\wedge C}^2 & + & nk\sigma_R^2 & + & nk\sigma_C^2 & + & n^2k\sigma_U^2.
\end{array}$$

Omitting $R \wedge C$ from the model is equivalent to assuming that $\xi_E = \xi_{R\wedge C}$. In terms of components of variance, this means that

$$\sigma_{R\wedge C}^2 = 0, \qquad (4)$$

so that, although rows and columns are represented by random variables, blocks are not. In terms of covariances it means that

$$\begin{aligned}
\gamma_{R\wedge C} &= \gamma_R + \gamma_C - \gamma_U \\
&= \sigma_R^2 + \sigma_C^2 + \sigma_U^2, \qquad (5)
\end{aligned}$$

so that the covariance between responses on two plots in the same block is determined by the covariances between pairs in the same row (but different blocks), those in the same column (but different blocks), and those in different rows and columns. In Examples 1, 2 and 4 assumptions (4) and (5) seem unreasonable to me. As a general rule then, I use model (1).

Table 1. Analysis-of-variance table if $R \wedge C$ is ignored

Stratum	Source	Degrees of freedom	
Mean			1
Rows			$n-1$
Columns			$n-1$
Plots	Treatments	$nk-1$	
	Residual	$(nk-2)(n-1)$	
			$n^2k-2n+1$
			n^2k

Example 3 is rather different, because the columns also represent treatments. It is argued by Preece et al. (1978) that a suitable model in such a case is to have an expectation part which is additive in C and S and covariance matrix

$$\mathcal{C} = \gamma_E I + \gamma_R (J_R - I) + \gamma_U (J - J_R). \tag{6}$$

The argument depends on the simultaneous randomization of both years' experiments. Once the relationship between R and C is fixed, there can be no randomization validity for an analysis which ignores $R \wedge C$. For this reason, Yates (1935) advised against such an analysis. Rojas and White (1957) studied the behaviour of such an analysis, using different randomizations of a semi-Latin square on uniformity data. They supported Yates' conclusion.

Recall that *strata* for the analysis of variance are just the eigenspaces of $\mathcal{C}$: see Nelder (1965) and Bailey (1981, 1991). If the model (6) without $R \wedge C$ is assumed, then the analysis of variance is as shown in Table 1, no matter what semi-Latin square is used. There is, therefore, no point in looking for anything more sophisticated than an orthogonal semi-Latin square.

If model (1) is assumed and an orthogonal semi-Latin square is used, then we obtain the analysis of variance in Table 2. There is now a separate stratum for $R \wedge C$, but the design is orthogonal in the sense that the treatment subspace is the direct sum of two parts, one in the $R \wedge C$ stratum and one in the plots stratum. The analysis is very similar to that for a classical split-plot design. If a non-orthogonal design is used then we effectively have the incomplete-block design $\Delta(\Omega)$, with the possibility of combining information from the bottom two strata. The difference from the design $\Delta(\Omega)$ is that the second stratum containing treatment information has dimension only $(n-1)^2$, not n^2-1. The analysis-of-variance table now depends on the semi-Latin square used. A Trojan square gives the analysis of variance in Table 3. Some statisticians may prefer to use only the information in the plots stratum. In this case the incomplete-block design $\Delta(\Omega)$ needs to be at least connected, and preferably optimal with respect to the usual optimality criteria such as A-optimality, D-optimality and E-optimality.

If $R \wedge C$ is included in the model as a *fixed* effect then the terms for R and C become redundant. The analysis is similar to that just given, except that there is no longer any possibility of recovering information which is orthogonal to the plots stratum. Now an orthogonal design is disastrous, because the contrasts between levels of L are totally confounded with blocks, and so are not estimable; the design $\Delta(\Omega)$ is not connected. It is even more important to use an optimal design in this case.

Table 2. Analysis-of-variance table for an orthogonal semi-Latin square if $R \wedge C$ is not ignored

Stratum	Source	Degrees of freedom	
Mean			1
Rows			$n-1$
Columns			$n-1$
Rows $\wedge$ Columns	L	$n-1$	
	Residual	$(n-1)(n-2)$	
			$(n-1)^2$
Plots	Treatments orthogonal to L	$n(k-1)$	
	Residual	$n(n-1)(k-1)$	
			$n^2(k-1)$
			n^2k

Table 3. Analysis-of-variance table for a Trojan square if $R \wedge C$ is not ignored

Stratum	Source	Degrees of freedom	
Mean			1
Rows			$n-1$
Columns			$n-1$
Rows $\wedge$ Columns	within squares (efficiency $1/k$)	$k(n-1)$	
	Residual	$(n-k-1)(n-1)$	
			$(n-1)^2$
Plots	within squares (efficiency $(k-1)/k$)	$k(n-1)$	
	between squares (efficiency 1)	$k-1$	
	Residual	$(nk-n-1)(n-1)$	
			$n^2(k-1)$
			n^2k

4. Randomization

The correct randomization for a semi-Latin square design whose rows and colmuns are inherent in the experimental material consists of the following three independent stages.

1. Randomize rows.
2. Randomize columns.
3. Randomize plots within each block independently.

Thus randomization could transform the design in Figure 5 into the plan in Figure 7.

Figure 7. Randomized verison of Figure 5

B	γ	d	b	α	E	c	β	A	C	δ	e	a	ε	D
C	b	ε	e	A	γ	δ	B	a	α	D	c	β	d	E
c	δ	E	β	C	a	D	γ	b	d	ε	A	B	e	α
a	A	α	D	δ	d	e	ε	E	B	b	β	C	γ	c
D	β	e	c	B	ε	C	α	d	γ	a	E	b	A	δ

Care is needed when the rows and columns are not geometric rows and columns. In Example 3, last year's varieties might be as shown in Figure 8, where the five replicates may be well separated and may be in different orientations. If there are 15 chemicals in the current year's experiment then the plan in Figure 7 must be interpreted as in Figure 9.

Figure 8. Plan for 5 varieties last year

4	3	1	5	2

2	1	4	3	5

1	3	5	2	4

4	3	2	5	1

2	5	4	1	3

Figure 9. Plan for 15 chemicals this year, following the plan in Figure 8

C δ e	c β A	B γ d	a ε D	b α E
e A γ	C b ε	α D c	δ B a	β d E
c δ E	D γ b	B e α	β C a	d ε A
B b β	e ε E	D δ d	C γ c	a A α
c B ε	b A δ	γ a E	D β e	C α d

In the sugar beet trials (Example 4) I find it best to randomize as in Figure 7 and then convert this to the field plan using the rule

ith plot in row j and column ℓ	$\longmapsto$	plot where drill ℓ crosses row j for the ith time.

Figure 7 thus gives the plan in Figure 10.

Figure 10. Plan in Figure 7 converted for sugar beet trials

Drill	1	2	3	4	5	5	4	3	2	1	1	2	3	4	5
	B	b	c	C	a	ε	δ	β	α	γ	d	E	A	e	D
	C	e	δ	α	β	d	D	B	A	b	ε	γ	a	c	E
	c	β	D	d	B	e	ε	γ	C	δ	E	a	b	A	α
	a	D	e	B	C	γ	b	ε	δ	A	α	d	E	β	c
	D	c	C	γ	b	A	a	α	B	β	e	ε	d	E	δ

5. Optimality of Trojan Squares

If $R \wedge C$ is included in the statistical model, then it is important to use a semi-Latin square Ω whose quotient incomplete-block design $\Delta(\Omega)$ has high efficiency factors. The efficiency factors for various semi-Latin squares were calculated by Bailey (1988). The following lemma gives an alternative derivation of the efficiency factors of Trojan squares and some other semi-Latin squares.

Lemma 5.1. *For $i = 1, \ldots, M$, let Δ_i be a binary incomplete-block design for a treatment set T_i in b blocks of size k_i, where $|T_i| = t_i$ and each treatment is replicated r times. Suppose that $T_i \cap T_j = \emptyset$ for $i \neq j$, and put $T = T_1 \cup \cdots \cup T_M$. Let $B_{i1}, \ldots, B_{ib}$ be the subsets of T_i which are in the blocks of Δ_i. Suppose that Δ is a binary incomplete-block design for treatment set T in b blocks of size k, where $k = \sum_i k_i$, such that*

(a) *the set of treatments in the* s*th block of* Δ *is* $B_{1s} \cup \cdots \cup B_{Ms}$

(b) *there are integers* λ_{ij} *for* $i \neq j$ *such that if* $\tau_i \in T_i$ *and* $\tau_j \in T_j$ *then the concurrence of* τ_i *and* τ_j *in* Δ *is* λ_{ij}.

Then the efficiency factors of Δ *are:*

$$\begin{array}{ll} 1-(1-e_i)\dfrac{k_i}{k} & \text{for each efficiency factor } e_i \text{ of } \Delta_i \\ 1 & \text{with (additional) multiplicity } M-1. \end{array}$$

Proof. Counting triples (τ_i, τ_j, B) with $\tau_i \in T_i$, $\tau_j \in T_j$ and block B containing both τ_i and τ_j, we obtain

$$t_i\, t_j\, \lambda_{ij} = bk_i\, k_j.$$

Since $bk_i = t_i r$ for all i, we obtain

$$\lambda_{ij} = r^2/b$$

for all i and j with $i \neq j$; and $\lambda_{ij} t_i = rk_i$.

Let N_i be the concurrence matrix of Δ_i. The concurrence matrix N of Δ is block diagonal with the N_i on the diagonal and all other entries equal to r^2/b. In $\mathbf{R}^{T_i}$, let u_i be the all-1s vector, let 0_i be the zero vector, and let v_i be an eigenvector of N_i with eigenvalue f_i, such that v_i is orthogonal to u_i. Then

$$v_i\left(I - \frac{1}{rk_i}N_i\right) = \left(1 - \frac{f_i}{rk_i}\right)v_i,$$

and so the canonical efficiency factor e_i associated with v_i is given by

$$e_i = 1 - \frac{f_i}{rk_i}.$$

In $\mathbf{R}^T$, put $w_i = (0_1, \ldots, 0_{i-1}, v_i, 0_{i+1}, \ldots, 0_M)$. Then

$$w_i\, N = f_i\, w_i\,,$$

so w_i is a canonical contrast for Δ with efficiency factor

$$1 - \frac{f_i}{rk} = 1 - (1-e_i)\frac{k_i}{k}.$$

Finally put $x_i = (0_1, \ldots, 0_{i-1}, u_i, 0_{i+1}, \ldots, 0_M)$ and $x_{ij} = k_i^{-1}x_i - k_j^{-1}x_j$. Then

$$x_i N = rk_i\, x_i + \sum_{j\neq i} \lambda_{ij}\, t_i\, x_j = rk_i\left(x_i + \sum_{j \neq i} x_j\right)$$

because $u_i N_i = r k_i u_i$. Thus,

$$x_{ij} N = 0,$$

and so x_{ij} is a canonical contrast for Δ with efficiency factor 1.

Corollary 5.2. *The efficiency factors of a $(n \times n)/k$ Trojan square are $1 - k^{-1}$, with multiplicity $k(n-1)$, and 1, with multiplicity $k-1$.*

Proof. In Lemma 5.1, let $\Delta_1, \ldots, \Delta_k$ be the constituent Latin squares of the Trojan square. Then $k_i = 1$ and all efficiency factors of Δ_i are zero.

Corollary 5.3. *Let $\Lambda_1, \ldots, \Lambda_s$ be mutually orthogonal $n \times n$ Latin squares. Let Δ_i be the semi-Latin square obtained by inflating Λ_i by k_i, and let Ω be the semi-Latin square obtained by superposing $\Delta_1, \ldots, \Delta_s$, so that Ω is an* OMA$(s+2, n; 1, 1, k_1, \ldots, k_s)$. *Let $k = \sum_i k_i$. Then the efficiency factors of Ω are $1 - k_i/k$, with multiplicity $n-1$, for $i = 1, \ldots, s$, and 1, with multiplicity $nk - ns + s - 1$.*

Proof. Now Δ_i has $n-1$ efficiency factors equal to zero and the rest equal to 1. Using these values in Lemma 5.1 gives the result.

Theorem 5.4. *If Ω is a Trojan square with $k = n-1$, or an inflation of such a square, then $\Delta(\Omega)$ is* A-, D- *and* E-*optimal among semi-Latin squares of that size.*

Proof. The dimension of the $R \wedge C$ stratum is equal to $(n-1)^2$, so at most $(n-1)^2$ of the efficiency factors can be less than 1, by the theory given by James and Wilkinson (1971). Corollaries 5.2 and 5.3 show that Ω has $nk - 1 - (n-1)^2$ efficiency factors equal to 1 and the remainder equal to $1 - 1/(n-1)$. Since the sum of the efficiency factors of a binary design is fixed, Ω maximizes the harmonic mean, the geometric mean, and the minimum of the efficiency factors.

Theorem 5.5. *If Ω is any Trojan square then $\Delta(\Omega)$ is* A-, D-, *and* E-*optimal among all binary incomplete-block designs of that size.*

Proof. See Cheng and Bailey (1991).

Note that the competing designs are different in Theorems 5.4 and 5.5. When $n = 3$ and $k = 4$ then Theorem 5.4 shows that a 2-fold inflation of the $(3 \times 3)/2$ Trojan square is optimal among semi-Latin squares. It has efficiency factors equal to $1/2$, with multiplicity 4, and 1, with multiplicity 7. However, among binary incomplete-block designs for 12 treatments in 9 blocks of size 4, the optimal design is the dual of the balanced lattice design. This has efficiency factors equal to $3/4$, with multiplicity 8, and 1, with multiplicity 3. By contrast, when $k < n$ then Theorem 5.5 shows that Trojan squares are optimal even over designs which are not constrained to be doubly-resolvable.

Definition. The integer pair (n, k) is *Trojan* if either

(a) $k < n$ and there exists a set of k mutually orthogonal $n \times n$ Latin squares, or

(b) k is a multiple of $n-1$ and there exists a set of $n-1$ mutually orthogonal $n \times n$ Latin squares.

Thus, if (n,k) is Trojan then we know how to find an optimal semi-Latin square: a Trojan square or an inflated $(n \times n)/(n-1)$ Trojan square. I do not know any optimality results for $(n \times n)/k$ semi-Latin squares if (n,k) is not Trojan, except when $n = 3$ (see Corollary 6.3).

Corollary 5.3 shows that Lemma 3, and hence part (b) of the Theorem, of Bailey (1988) are false. The latter result concerns the optimality of semi-Latin squares with $k \geq n$. I have never had to use such a square in practice. However, they cannot be ruled out. Indeed, a square with $n = k = 4$ has been used by Fielding (1990) for banana trees in Jamaica. The most likely values of k would be close to n, and so not covered by Theorem 5.4.

6. Other Constructions

There are many other semi-Latin squares in addition to those formed from Latin squares by inflation and superposition, or a combination thereof. Some constructions are given by Andersen (1979), Andersen and Hilton (1980a,b), Preece and Freeman (1983) and Bailey (1988). It might be possible to enumerate the isomorphism classes of $(n \times n)/k$ semi-Latin squares for small values of nk.

Theorem 6.1. *There is only one isomorphism class of $(2 \times 2)/k$ semi-Latin squares, for each k. It consists of the k-fold inflation of the 2×2 Latin square.*

Theorem 6.2. *There are $\lfloor k/2 \rfloor + 1$ isomorphism classes of $(3 \times 3)/k$ semi-Latin squares.*

Proof. Let B_{ij} be the block in the ith row and jth column. Let T_1 be the set of symbols common to B_{11} and B_{22}, T_2 the set of symbols common to B_{11} and B_{23}, T_3 the remaining symbols in B_{22}, and T_4 the remaining symbols in B_{23}. We must have

$T_1 \ T_2$	T_4	T_3
	$T_1 \ T_3$	$T_2 \ T_4$
$T_3 \ T_4$	T_2	T_1

,

which can be completed only as

$T_1 \ T_2$	$T_4 \ T_5$	$T_3 \ T_6$
$T_5 \ T_6$	$T_1 \ T_3$	$T_2 \ T_4$
$T_3 \ T_4$	$T_2 \ T_6$	$T_1 \ T_5$

,

where $|T_1| = |T_4| = |T_6| = k_1$, say, and $|T_2| = |T_3| = |T_5| = k - k_1$. Thus the semi-Latin square is the superposition of the k_1-fold inflation of a 3×3 Latin square Λ_1 with the $(k - k_1)$-fold inflation of the 3×3 Latin square Λ_2 orthogonal to Λ_1. Interchanging Λ_1 with Λ_2 does not change the isomorphism class, which is therefore determined by $\min\{k_1, k - k_1\}$.

Corollary 6.3. *If $k = 2s + 1$ then the superposition of an s-fold inflation of a 3×3 Latin square Λ_1 with an $(s + 1)$-fold inflation of a 3×3 Latin square Λ_2 orthogonal to Λ_1 is an optimal $(3 \times 3)/k$ semi-Latin square.*

Preece and Freeman (1983) list eleven isomorphism classes of $(4 \times 4)/2$ semi-Latin squares, and claim that there are no more. It should not be too difficult to enumerate the isomorphism classes of $(4 \times 4)/k$ semi-Latin squares for general k, by allocating non-negative integers to each of the 24 transversals of the 4×4 array and using the constraints that the sum of these integers on any block is equal to k.

In view of the results in Section 5, practical use of semi-Latin squares requires us to find further constructions only for non-Trojan pairs (n, k). The total number of isomorphism classes for such a pair may be too large to investigate completely. However, it is widely conjectured that optimal designs are to be found among the regular-graph designs—those in which no two concurrences differ by more than 1: see John and Mitchell (1977), John and Williams (1982). Thus the search for efficient semi-Latin squares for non-Trojan (n, k) could reasonably be restricted to squares Ω for which $\Delta(\Omega)$ is a regular-graph design, if any such exist.

Theorem 6.4. *Let Ω be an $(n \times n)/k$ semi-Latin square such that $\Delta(\Omega)$ is a regular-graph design. Then*

(i) *no concurrence in $\Delta(\Omega)$ is greater than 1;*
(ii) *$k \leq n - 1$;*
(iii) *if $k = n - 1$ then Ω is Trojan.*

Proof.

(i) The average concurrence is equal to $n(k-1)/(nk-1)$, which is less than 1.
(ii) The k symbols in the block in the first row and first column must all be in different blocks in the second row, none of which is the block in the first column.
(iii) Let T be the set of symbols in a given block B. Because $|T| = n - 1$, each block B' in different rows and columns from those containing B contains a single element of T. Let τ be a symbol outside T. Then τ occurs in $n - 2$ such blocks B', and so concurs with $n - 2$ elements of T. Thus if any two treatments concur then any third treatment must concur with at least one

of them. It follows that the relation 'is equal to or does not concur with' on the set of symbols is an equivalence relation with classes of size n. Thus, $\mathcal{G}(\Omega)$ is the complete k-partite graph on sets of size n, and so Ω is Trojan.

Theorem 6.4 gives no guidance on optimal semi-Latin squares if $k \geq n$. The following conjecture is guided by Theorem 5.4 and Corollary 6.3.

Conjecture 6.5. *If $\Lambda_1, \ldots, \Lambda_{n-1}$ is a set of mutually orthogonal $n \times n$ Latin squares and $k = a(n-1) + b$ with $a \geq 1$ and $1 \leq b < n-1$, and Ω is the superposition of the $(a+1)$-fold inflations of $\Lambda_1, \ldots, \Lambda_b$ with the a-fold inflations of $\Lambda_{b+1}, \ldots, \Lambda_{n-1}$, then Ω is an optimal semi-Latin square.*

Conjecture 6.5 would be a useful result if true, but it seems counter to intuition because it recommends semi-Latin squares some of whose concurrences λ_{ij} are as large as n. In their important work on singly resolvable block designs Patterson and Williams (1976b) recommend using designs with $\lambda_{ij} \in \{0, 1, 2\}$ in these circumstances, while Paterson (1983) recommends using designs which minimize the number of short circuits in the variety-concurrence graph, in particular $\sum_{i \neq j} \lambda_{ij}^2$.

Example 8. Compare the two $(4 \times 4)/4$ semi-Latin squares in Figure 11. The square Ω_1 in Figure 11(a) is the superposition of three mutually orthogonal Latin squares Λ_1, Λ_2, Λ_3 (on large Latin, small Greek and small Latin alphabets respectively) with a fourth Latin square Λ_4 (on the digits $1 \ldots 4$) which may be obtained from any of $\Lambda_1, \Lambda_2, \Lambda_3$ by transposing a pair of rows. The square Ω_2 in Figure 11(b) is the superposition of the 2-fold inflation of Λ_1 with Λ_2 and Λ_3. The concurrences of Ω_1 are 0 (48 times), 1 (48 times) and 2 (24 times), while those of Ω_2 are 0 (36 times), 1 (80 times) and 4 (4 times). Thus Patterson and Williams' advice would be to use Ω_1, while Conjecture 6.5 favours Ω_2. Both squares have $\sum \lambda_{ij}^2 = 144$, and $\mathcal{G}(\Omega_1)$ and $\mathcal{G}(\Omega_2)$ both contain 256 triangles, so Paterson's conjecture does not distinguish between Ω_1 and Ω_2 until quadrilaterals are counted. Corollary 5.3 shows that the efficiency factors of Ω_2 are

$$
\begin{array}{ll}
1 & \text{with multiplicity } 6 \\
\frac{3}{4} & \text{with multiplicity } 6 \\
\frac{1}{2} & \text{with multiplicity } 3,
\end{array}
$$

while the symmetry calculations of Bailey and Rowley (1990) show that these are also the efficiency factors of Ω_1. The two semi-Latin squares are therefore equally good according to all the usual optimality criteria. Whether or not Conjecture

6.5 is true, this suggests that intuition may be a poor guide to optimality for $k \geq n$.

Definition. Semi-Latin squares Ω_1 and Ω_2 of size $(n \times n)/k_1$ and $(n \times n)/k_2$ are *orthogonal* to each other if $\Delta(\Omega_1)$ and $\Delta(\Omega_2)$ satisfy the hypotheses of Lemma 5.1 with the blocks in corresponding order.

Conjecture 6.6. *If Ω_1 and Ω_2 are optimal semi-Latin squares and they are orthogonal to each other then their superposition is also optimal.*

Figure 11. Two semi-Latin squares with $n = k = 4$

$A\ \alpha\ a\ 1$	$B\ \beta\ b\ 2$	$C\ \gamma\ c\ 3$	$D\ \delta\ d\ 4$
$B\ \gamma\ d\ 2$	$A\ \delta\ c\ 1$	$D\ \alpha\ b\ 4$	$C\ \beta\ a\ 3$
$C\ \delta\ b\ 4$	$D\ \gamma\ a\ 3$	$A\ \beta\ d\ 2$	$B\ \alpha\ c\ 1$
$D\ \beta\ c\ 3$	$C\ \alpha\ d\ 4$	$B\ \delta\ a\ 1$	$A\ \gamma\ b\ 2$

(a) The semi-Latin square Ω_1

$A_1\ A_2\ \alpha\ a$	$B_1\ B_2\ \beta\ b$	$C_1\ C_2\ \gamma\ c$	$D_1\ D_2\ \delta\ d$
$B_1\ B_2\ \gamma\ d$	$A_1\ A_2\ \delta\ c$	$D_1\ D_2\ \alpha\ b$	$C_1\ C_2\ \beta\ a$
$C_1\ C_2\ \delta\ b$	$D_1\ D_2\ \gamma\ a$	$A_1\ A_2\ \beta\ d$	$B_1\ B_2\ \alpha\ c$
$D_1\ D_2\ \beta\ c$	$C_1\ C_2\ \alpha\ d$	$B_1\ B_2\ \delta\ a$	$A_1\ A_2\ \gamma\ b$

(b) The semi-Latin square Ω_2

7. Efficient Designs for non-Trojan Pairs

Real experiments are no respecters of pure mathematics. They may well need efficient semi-Latin squares for non-Trojan pairs (n, k). In the sugar beet trials (Example 4) at Brooms Barn Experimental Station in England, the number n of drills is usually equal to 6: see Bailey and Payne (1990). As (6, k) is not Trojan for any k except $k = 1$, we therefore need to find efficient $(6 \times 6)/k$ semi-Latin squares. So far, our search for such designs has been confined to regular-graph designs. Theorem 6.4 shows that these do not exist unless $k \leq 4$.

An efficient $(6 \times 6)/2$ semi-Latin square was found independently by Brickell (1984) and Bailey (1990). It is shown in Figure 12. Its efficiency factors are:

$$\frac{1}{2} = 0.5000 \quad \text{with multiplicity } 5$$

$$\frac{7+\sqrt{5}}{12} = 0.7697 \quad \text{with multiplicity } 3$$

$$\frac{7-\sqrt{5}}{12} = 0.3970 \quad \text{with multiplicity } 3.$$

Table 4 compares various measures of overall efficiency with those that a Trojan square would have, if it existed. The minimum efficiency factor, the harmonic mean and the geometric mean of the efficiency factors are the E-, A- and D-criteria respectively. The minimum simple and maximum simple efficiency factors are the extreme values of the efficiency factors of contrasts which compare one treatment with another. Of course, the quotient block design of a Trojan square is a connected group divisible design with k groups of size n and concurrences equal to 0 and 1. When $k = 2$, such a design always exists and the Corollary to Theorem 3.1 of Cheng (1978) shows that it is optimal. Thus, the competing design in Table 4 is optimal for 36 blocks of size 2 even though it is not in the list of optimal designs given by John and Mitchell (1977) and does not exist as a semi-Latin square.

Figure 12. Efficient $(6 \times 6)/2$ semi-Latin square

A L	*F K*	*B G*	*C H*	*D I*	*E J*
C I	*B J*	*H L*	*E F*	*G K*	*A D*
D J	*A E*	*C K*	*I L*	*B F*	*G H*
E K	*H I*	*A F*	*D G*	*J L*	*B C*
F G	*C D*	*I J*	*A B*	*E H*	*K L*
B H	*G L*	*D E*	*J K*	*A C*	*F I*

Table 4. Comparison of $(6 \times 6)/2$ semi-Latin squares

efficiency factors	Figure 12	hypothetical Trojan square
minimum	0.3970	0.5000
minimum simple	0.4783	0.5000
harmonic mean	0.5127	0.5238
geometric mean	0.5281	0.5325
maximum simple	0.5500	0.5455
maximum	0.7697	1.0000

A $(6 \times 6)/3$ semi-Latin square was needed for one of the 1990 trials at Brooms Barn. Peter Wild of Royal Holloway and Bedford New College, University of London, found a regular-graph-design semi-Latin square just in time for the trial. It is the superposition of the semi-Latin square in Figure 12 with the Latin square in Figure 13. Since these two are orthogonal to each other, Lemma 5.1 shows that the efficiency factors for the superimposed design are:

$$\frac{2}{3} = 0.6667 \quad \text{with multiplicity } 10$$

$$\frac{13+\sqrt{5}}{18} = 0.8464 \quad \text{with multiplicity } 3$$

$$\frac{13-\sqrt{5}}{18} = 0.5980 \quad \text{with multiplicity } 3$$

$$1 \quad \text{with multiplicity } 1.$$

In fact, this $(6 \times 6)/3$ semi-Latin square was given, as an OMA(4,6; 1, 1, 1, 2) by Brickell (1984), who found it by a computer search. Peter Wild's invaluable contribution was to realise that this example from communications theory was precisely what was needed to solve the statistical problem.

Figure 13. Latin square orthogonal to the semi-Latin square in Figure 12

1	2	3	5	6	4
2	1	6	3	4	5
3	6	1	4	5	2
5	3	4	1	2	6
6	4	5	2	1	3
4	5	2	6	3	1

Table 5. Comparison of $(6 \times 6)/3$ semi-Latin squares

efficiency factors	Figures 12 and 13	hypothetical Trojan square
minimum	0.5980	0.6667
minimum simple	0.6586	0.6667
harmonic mean	0.6922	0.6939
geometric mean	0.6986	0.6992
maximum simple	0.7099	0.7059
maximum	1.0000	1.0000

Table 5 compares this semi-Latin square with a hypothetical Trojan square. The competing design exists as the dual of a three-replicate square lattice design, which is optimal among designs with 36 blocks of size 3, by Corollary 2.3 of Cheng and Bailey (1991).

References

Andersen, L. D. (1979). Latin squares and their generalizations. Ph.D. thesis, University of Reading.

Andersen, L. D. and Hilton, A. J. W. (1980a). Generalized Latin rectangles. I: Construction and decomposition. *Discrete Math.* **31**, 125–152.

Andersen, L. D. and Hilton, A. J. W. (1980b). Generalized Latin rectangles. II: Embedding. *Discrete Math.* **31**, 235–260.

Anthony, M. H. G., Martin, K. M., Seberry, J. and Wild, P. (1990). Some remarks on authentication systems. In *Advances in Cryptology, Auscrypt '90, Lecture Notes in Comput. Sci.* **453**, 122–139. Springer-Verlag, New York.

Bailey, R. A. (1981). A unified approach to design of experiments. *J. Roy. Statist. Soc. Ser. A* **144**, 214–223.

Bailey, R. A. (1984). Contribution to the discussion of "Analysis of variance models in orthogonal designs" by T. Tjur. *Internat. Statist. Rev.* **52**, 65–77.

Bailey, R. A. (1985). Factorial design and Abelian groups. *Linear Algebra Appl.* **70**, 349–368.

Bailey, R. A. (1988). Semi-Latin squares. *J. Statist. Plann. Infer.* **18**, 299–312.

Bailey, R. A. (1989). Designs: mappings between structured sets. In *Surveys in Combinatorics, 1989* (Edited by J. Siemons), London Mathematical Society Lecture Notes, **141**, 22–51. Cambridge University Press, Cambridge.

Bailey, R. A. (1990). An efficient semi-Latin square for twelve treatments in blocks of size two. *J. Statist. Plann. Infer.* **26**, 263–266.

Bailey, R. A. (1991). Strata for randomized experiments. *J. Roy. Statist. Soc. Ser. B* **53**, 27–66.

Bailey, R. A. and Payne, R. W. (1990). Experimental design: statistical research and its application. In *Institute of Arable Crops Research Report for 1989* (Edited by J. Abbott), 107–112. Agriculture and Food Research Council, Institute of Arable Crops Research, Harpenden.

Bailey, R. A. and Rowley, C. A. (1990). General balance and treatment permutations. *Linear Algebra Appl.* **127**, 183–225.

Bailey, R. A. and Speed, T. P. (1986). Rectangular lattice designs: efficiency factors and analysis. *Ann. Statist.* **14**, 874–895.

Brickell, E. F. (1984). A few results in message authentication. *Congressus Numerantium* **43**, 141–154.

Cheng, C.-S. (1978). Optimality of certain asymmetrical experimental designs. *Ann. Statist.* **6**, 1239–1261.

Cheng, C.-S. and Bailey, R. A. (1991). Optimality of some two-associate-class partially balanced incomplete-block designs. *Ann. Statist.* **19**, 1667–1671.

Darby, L. A. and Gilbert, N. (1958). The Trojan square. *Euphytica* **7**, 183–188.

Fielding, W. (1990). Personal communication.

Harshbarger, B. and Davis, L. L. (1952). Latinized rectangular lattices. *Biometrics* **8**, 73–84.

James, A. T. and Wilkinson, G. N. (1971). Factorization of the residual operator and canonical decomposition of nonorthogonal factors in the analysis of variance. *Biometrika* **58**, 279–294.

John, J. A. and Mitchell, T. J. (1977). Optimal incomplete block designs. *J. Roy. Statist. Soc. Ser. B* **39**, 39–43.

John, J. A. and Williams, E. R. (1982). Conjectures for optimal block designs. *J. Roy. Statist. Soc. Ser. B* **44**, 221–225.

Nelder, J. A. (1965). The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance. *Proc. Roy. Soc. Ser. A* **283**, 147–162.

Paterson, L. (1983). Circuits and efficiency in incomplete block designs. *Biometrika* **70**, 215–225.

Patterson, H. D. and Williams, E. R. (1976a). Some theoretical results on general block designs. *Congressus Numerantium* **15**, 489–496.

Patterson, H. D. and Williams, E. R. (1976b). A new class of resolvable incomplete block designs. *Biometrika* **63**, 83–92.

Preece, D. A., Bailey, R. A. and Patterson, H. D. (1978). A randomization problem in forming designs with superimposed treatments. *Austral. J. Statist.* **20**, 111–125.

Preece, D. A. and Freeman, G. H. (1983). Semi-Latin squares and related designs. *J. Roy. Statist. Soc. Ser. B* **45**, 267–277.

Rasch, D. and Herrendörfer, G. (1986). *Experimental Design.* Reidel, Dordrecht.

Rojas, B. and White, R. F. (1957). The modified Latin square. *J. Roy. Statist. Soc. Ser.B* **19**, 305–317.

Speed, T. P. (1987). What is an analysis of variance? *Ann. Statist.* **15**, 885–910.

Speed, T. P. and Bailey, R. A. (1987). Factorial dispersion models. *Internat. Statist. Rev.* **55**, 261–277.

Thompson, R. (1984). The use of multiple copies of data in forming and interpreting analysis of variance. In *Experimental Design, Statistical Models, and Genetic Statistics. Essays in Honor of Oscar Kempthorne* (Edited by K. Hinklemann), 155–173. Marcel Dekker, New York.

Tjur, T. (1984). Analysis of variance models in orthogonal designs. *Internat. Statist. Rev.* **52**, 33–65.

Williams, E. R. (1986). Row and column designs with contiguous replicates. *Austral. J. Statist.* **28**, 154–163.

Yates, F. (1935). Complex experiments. *J. Roy. Statist. Soc. Suppl.* **2**, 181–247.

Department of Math. Studies, Goldsmiths' College, University of London, New Cross, London SE14 6NW, U.K.

(Received April 1991; accepted January 1992)