Determining the Parametric Structure of Models

D.J. Cole\textsuperscript{a}\textsuperscript{*}, B.J.T. Morgan\textsuperscript{a}, D.M. Titterington\textsuperscript{b}

\textsuperscript{a}School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, CT2 7NF, UK
\textsuperscript{b}Department of Statistics, University of Glasgow, Glasgow G12 8QQ, UK

Abstract

In this paper we develop a comprehensive approach to determining the parametric structure of models. This involves considering whether a model is parameter redundant or not and investigating model identifiability. The approach adopted makes use of exhaustive summaries, quantities that uniquely define the model. We review and generalise previous work on evaluating the symbolic rank of an appropriate derivative matrix to detect parameter redundancy, and then develop further tools for use within this framework, based on a matrix decomposition. Complex models, where the symbolic rank is difficult to calculate, may be simplified structurally using reparameterisation and by finding a reduced-form exhaustive summary. The approach of the paper is illustrated using examples from ecology, compartment modelling and Bayes networks. This work is topical as models in the biosciences and elsewhere are becoming increasingly complex.

\textbf{Keywords:} derivative matrix, exhaustive summaries, global identifiability, Jacobian matrix, local identifiability, parameter redundancy

1. Introduction

1.1. Aims and outline of the paper

Statistical inference may fail due to an inability to estimate, or estimate well, all of the parameters of a model. This may be because of a lack of data. However,
for some models it will never be possible to estimate all of the parameters by methods of classical inference. For example this can occur if two parameters are confounded and only ever appear as a product. In such instances a model is termed non-identifiable or parameter redundant.

For linear models we can use a constant design matrix to determine parameter redundancy and impose appropriate constraints. For non-linear models, the subject of this paper, it can often be difficult to determine if a model is parameter redundant. One approach is to use a symbolic algebra computer package, which involves forming a suitable derivative matrix and then calculating its symbolic rank. The idea of determining parameter redundancy through the rank of a derivative matrix, which corresponds to the use of the design matrix for linear models, appears in [1, 2, 3, 4, 5] and others. However, the early work predated symbolic algebra packages. A more detailed discussion of previous work is given in Section 1.3. The early papers differed in what was differentiated with respect to the parameters to form the derivative matrix. Here we provide a unifying framework by first defining the quantity that is differentiated as an exhaustive summary. This is a term borrowed from compartment modelling [6], and is a vector of parameter combinations that uniquely defines the structure of the model. A full discussion of exhaustive summaries and how they can be used to determine the parametric structure of models is given in Section 2.

The symbolic approach is explained in Section 2.1. However, for structurally complex models, computer algebra packages may not be able to calculate the symbolic rank of the derivative matrices, due to computer memory limitations. In such cases, numerical methods have been used. To overcome this difficulty we need to use an exhaustive summary that is structurally simple. Here we present a collection of tools that can be used to create such exhaustive summaries. The most powerful tool is reparameterisation of the model in order to find a new simpler exhaustive summary based on the reparameterisation, and this is the subject of Section 3. We also employ the extension theorems of [5, 7] which allows conclusions regarding specific cases of a model to be extended to more general forms of the same model structure (Section 2.2).
We present the results of the paper through a formal presentation of theorems, remarks and examples, complemented by reference to Maple work sheets. The tools of this paper provide a framework that can be used to determine the parametric structure of models. If a model is parameter redundant, exhaustive summaries can be used to determine what combination of parameters can be estimated. If the model is not parameter redundant it is termed full rank, and exhaustive summaries can be used to show if the model is always full rank or if there are points in the parameter space where the model is parameter redundant (Section 2.3) or if a model is only full rank in a region in the parameter space (Section 3.1). Several simple examples are given alongside the development of the theory, and extended examples are presented in Section 4. Note that the first 3 examples are simple examples where existing theory is sufficient to examine whether the model is parameter redundant; these examples are used to illustrate clearly the existing and new methodology. The later examples show how the approach of the paper provides a general method that may be used in many different areas, and is often a viable alternative to developing a specific method for a particular class of problems. Implications for both classical and Bayesian inference are discussed in Section 5. The Maple code for the examples can be found at www.kent.ac.uk/ims/personal/djc24/maplecode.htm.

Partly as a result of fast computers, we are seeing an increase in the complexity of models being used throughout biosciences. There is therefore a current need for the developments in this paper. For instance [8, 9] provide examples from capture-recapture and latent class modelling where the methods of this paper provide definitive answers to essential questions of parameter redundancy which were not known previously and only examined through time-consuming numerical investigations.

1.2. Definitions of Identifiability and Parameter Redundancy

A model is not identifiable if different sets of parameter values result in the same model [10]. More formally, let $M(\theta)$ be the function that defines a model, which has unknown parameters $\theta \in \Omega$, where $\Omega$ is a $\text{dim}(\theta)$-dimensional
vector space and dim(θ) denotes the number of terms in a general vector θ. For example \( M(\theta) \) could be a suitable probability distribution.

**Definition 1.** A model is globally identifiable if \( M(\theta_1) = M(\theta_2) \) implies that \( \theta_1 = \theta_2 \). A model is locally identifiable if there exists an open neighbourhood of any \( \theta \) such that this is true. Otherwise a model is non-identifiable.

Non-identifiability occurs if a model has too many parameters, which is termed parameter redundancy. A parameter redundant model can be written in terms of a smaller set of parameters [5].

**Definition 2.** A model is parameter redundant if we can write \( M(\theta) \) as a function just of \( \beta \), where \( \beta = f(\theta) \in \Omega_\beta \), in which \( \Omega_\beta \) has dimension \( \text{dim}(\beta) < \text{dim}(\theta) \).

Models which are not parameter redundant are described as full rank.

**Definition 3.** An essentially full rank model is full rank for all \( \theta \). A conditionally full rank model is not full rank for all \( \theta \).

This distinction depends upon the specification of the parameter space \( \Omega \).

### 1.3. Previous use of symbolic methods

#### 1.3.1. Symbolic rank of a derivative matrix.

Models where \( M(\theta) \) is the exponential family probability density function were considered in [5], who showed that whether or not a model defined by \( M(\theta) \) is parameter redundant can be determined by checking the symbolic rank of the derivative matrix, \( D = [\partial \mu_k / \partial \theta_i] \), where \( \mu_k \) is the expectation of the \( k \)th observation, \( y_k \), and \( \theta_i \) is the \( i \)th model parameter. If the symbolic rank of \( D \) is equal to the number of parameters \( p \), the number of rows of \( D \), then the model is full rank. If the symbolic rank of \( D \) is less than \( p \) the model is parameter redundant and not identifiable. The symbolic rank of \( D \) can in principle be obtained by using a symbolic algebra computer package such as Maple; see [11]. In earlier work on exponential family models, [12] adopted
the equivalent criterion of differentiating the canonical parameter. We shall use derivative matrix terminology, whereas others sometimes refer to such matrices as Jacobians.

**Example 1 - The Cormack Jolly Seber model for capture-recapture data.**

This method can be illustrated using the Cormack Jolly Seber (CJS) model [13, 14, 15]. This is a product-multinomial model used in analysing capture-recapture data in ecology. In this model \( N_i \) animals are marked in year \( i \), for \( i = 1, \ldots, r \), and then it is recorded when they are next seen alive. All of the model parameters are time-dependent. The probability that an animal marked in year \( i \) is first seen in year \( j \) (for \( i < j = 1, \ldots, c \)) is

\[
p_{i,j} = \left( \prod_{k=i}^{j-1} \phi_k \right) \left( \prod_{k=i+1}^{j} \bar{p}_k \right) p_{j+1},
\]

where \( \phi_{k+1} \) is the probability of surviving year \( k \), \( p_k \) is the probability of recapture in year \( k \) and \( \bar{p}_k = 1 - p_k \). The probability that an animal is never seen again is

\[
1 - \sum_{j=i}^{c} p_{i,j}, \text{ for } 1 \leq i \leq r.
\]

The mean \( \mu \) of this product-multinomial model is then made up of terms of the form \( N_i p_{i,j} \) and \( N_i(1 - \sum_{j=i}^{c} p_{i,j}) \). Consider the model with 3 years of marking and 3 years of recapture \( (r = c = 3) \); the mean is then

\[
\mu = \begin{bmatrix}
N_1\phi_1 p_2 \\
N_1\phi_1 \phi_2 p_2 p_3 \\
N_1\phi_1 \phi_2 \phi_3 \bar{p}_2 p_3 p_4 \\
N_1(1 - \phi_1 p_2 - \phi_1 \phi_2 p_2 p_3 - \phi_1 \phi_2 \phi_3 \bar{p}_2 p_3 p_4) \\
N_2\phi_2 p_3 \\
N_2\phi_2 \phi_3 \bar{p}_3 p_4 \\
N_2(1 - \phi_2 p_3 - \phi_2 \phi_3 \bar{p}_3 p_4) \\
N_3\phi_3 p_4 \\
N_3(1 - \phi_3 p_4)
\end{bmatrix}, \tag{1}
\]

and the parameters are \( \theta = \begin{bmatrix} \phi_1 \phi_2 \phi_3 p_2 p_3 p_4 \end{bmatrix} \). As a result of differentiating the elements of \( \mu \) with respect to the elements of \( \theta \), a derivative matrix is formed as given in Table 1. It is also shown in [5] that it is sufficient to consider a derivative matrix formed from differentiating the non-zero \( p_{i,j} \) or
\ln(p_{i,j}) with respect to \(\theta\). Derivative matrices can be simplified further by multiplying by any elementary matrix as this does not change the derivative matrix rank. For example [7] suggest the scaled derivative matrix \(\text{diag}(\theta)D\). Some of these alternative derivative matrices are also given in Table 1. All of the derivative matrices can be seen to have symbolic rank 5, but differ in complexity. The conclusion is that as there are six parameters the model is parameter redundant. Here this is obvious because the parameters \(\phi_3\) and \(p_4\) only ever appear as the product, \(\phi_3 p_4\), and are therefore confounded. □

The idea of using a derivative matrix and its rank to test for the identifiability of a model, rather than parameter redundancy, was considered for a general econometric model by [1]. He showed that if the expected information matrix, \(I = -E \left[(\partial^2 \log f)/(\partial \theta \partial \theta^t)\right]\), for probability density function \(f\), is non-singular then the model is locally identifiable. In [5, 12] this test is shown to be equivalent to finding the rank of the derivative matrix for exponential family models. [1] also shows how ‘reduced-form’ parameters can be used to determine identifiability. Reduced-form parameters, \(h_i(\theta)\), are functions of the original parameters and, if the model is rewritten in terms of the \(h_i\), then all the \(h_i\) are identifiable. If the derivative matrix \[\partial h_i/\partial \theta_j\] has rank \(p\), then \(\theta\) was shown to be locally identifiable. Further, if the functions \(h_i\) are linear then conditions are given for when \(\theta\) is globally identifiable.

Other users of derivative matrices and their ranks to determine model parametric structure include [2] for latent-class models, [4] for non-linear regression models and a range of authors within the area of compartment modelling and dynamic systems, which are considered below.

For any linear compartment model, \(M(\theta)\) is the output function:

\[y(t, \theta) = Cx(t, \theta), \quad \text{with} \quad \frac{\partial}{\partial t}x(t, \theta) = A(\theta)x(t, \theta) + B(\theta)u(t),\]

for suitable matrices \(A, B\) and \(C\), and input function \(u\). The transfer function approach to determine parametric structure in compartment models was introduced by [16]. This involves taking Laplace transforms to form a transfer function \(Q(s) = \tilde{y}(s)/\tilde{u}(s)\), where \(\tilde{y}(s)\) is the Laplace transform of \(y(t)\), etc.
with respect to the parameters to form the derivative matrix. The last case includes the scaled derivative matrix, \( \text{diag}(\theta)D \), as suggested in [7].

<table>
<thead>
<tr>
<th>( \kappa )</th>
<th>Derivative Matrix</th>
</tr>
</thead>
</table>
| \( \mu \) | \( D_{1.6.1...4} = \)
| 0 \( N_1p_2 \) | \( N_1\phi_2p_2p_3 \) | \( N_1\phi_2\phi_3p_2p_3p_4 \) | \( -N_1(p_2 + \phi_2p_2(p_3 + \phi_3p_3p_4)) \) |
| 0 | \( N_1\phi_1p_2p_3 \) | \( N_1\phi_1\phi_2p_3p_4 \) | \( -N_1\phi_1p_2(p_3 + \phi_3p_3p_4) \) |
| 0 | 0 | \( N_1\phi_1\phi_2p_2p_3p_4 \) | \( -N_1\phi_1\phi_2p_2p_3p_4 \) |
| \( N_1\phi_1 \) | \( -N_1\phi_1\phi_2p_2p_3 \) | \( -N_1\phi_1\phi_2\phi_3p_2p_3p_4 \) | \( -N_1\phi_1(1 - p_2(\phi_2 + \phi_2\phi_3p_4)) \) |
| 0 | \( N_1\phi_1\phi_2p_2 \) | \( -N_1\phi_1\phi_2\phi_3p_2p_4 \) | \( -N_1\phi_1p_2(\phi_2 - \phi_2\phi_3p_4) \) |
| 0 | 0 | \( N_1\phi_1\phi_2\phi_3p_2p_3 \) | \( -N_1\phi_1\phi_2\phi_3p_2p_3 \) |

| \( p_{ij} \) | \( D = \)
| 0 | \( p_2 \) | \( \phi_2p_2p_3 \) | \( \phi_2\phi_3p_2p_3p_4 \) | 0 | 0 | 0 |
| 0 | \( \phi_1p_2p_3 \) | \( \phi_1\phi_2p_2p_3p_4 \) | \( p_3 \) | \( p_3\phi_3p_4 \) | 0 |
| 0 | 0 | \( \phi_1\phi_2p_2p_3p_4 \) | 0 | \( \phi_2p_3p_4 \) | \( p_4 \) |
| \( \phi_1 \) | \( -\phi_1\phi_2p_3 \) | \( -\phi_1\phi_2\phi_3p_2p_3p_4 \) | 0 | 0 | 0 |
| 0 | \( \phi_1\phi_2p_2 \) | \( -\phi_1\phi_2\phi_3p_2p_4 \) | \( \phi_2 \) | \( -\phi_2\phi_3p_4 \) | 0 |
| 0 | 0 | \( \phi_1\phi_2\phi_3p_2p_3 \) | 0 | \( \phi_2\phi_3p_3 \) | \( \phi_3 \) |

| \( \ln(p_{ij}) \) | \( D = \)
| \( \phi_1^{-1} \) | \( \phi_1^{-1} \) | \( \phi_1^{-1} \) | 0 | 0 | 0 |
| 0 | \( \phi_2^{-1} \) | \( \phi_2^{-1} \) | \( \phi_2^{-1} \) | \( \phi_2^{-1} \) | 0 |
| 0 | 0 | \( \phi_3^{-1} \) | \( \phi_3^{-1} \) | \( \phi_3^{-1} \) | \( \phi_3^{-1} \) |
| \( p_2^{-1} \) | \( -\tilde{p}_2^{-1} \) | \( -\tilde{p}_2^{-1} \) | 0 | 0 | 0 |
| 0 | \( p_3^{-1} \) | \( -\tilde{p}_3^{-1} \) | \( p_3^{-1} \) | \( -\tilde{p}_3^{-1} \) | 0 |
| 0 | 0 | \( p_4^{-1} \) | \( p_4^{-1} \) | \( p_4^{-1} \) | \( p_4^{-1} \) |

| \( \text{diag}(\theta)D = \)
| 1 | 1 | 1 | 0 | 0 | 0 |
| 0 | 1 | 1 | 1 | 1 | 0 |
| 0 | 0 | 1 | 0 | 1 | 1 |
| 1 | \( -p_2\tilde{p}_2^{-1} \) | \( -p_2\tilde{p}_2^{-1} \) | 0 | 0 | 0 |
| 0 | 1 | \( -p_3\tilde{p}_3^{-1} \) | \( 1 - p_3\tilde{p}_3^{-1} \) | 0 |
| 0 | 0 | 1 | 0 | 1 | 1 |

Table 1: Various derivative matrices for the CJS model, all of rank 5. \( \kappa \) is what is differentiated with respect to the parameters to form the derivative matrix. The last case includes the scaled derivative matrix, \( \text{diag}(\theta)D \), as suggested in [7].
The numerator and denominator of $Q(s)$ are both polynomials in $s$ and the non-constant coefficients of the powers of $s$ are set equal to constants $\kappa_i$; these equations are called the moment invariants. This results in a set of equations that can be solved to find $\theta$ in terms of $\kappa$. If there is only one solution then the model is globally identifiable; if there are a countable number of solutions, the model is locally identifiable; if there are an infinite number of solutions the model is not identifiable [17]. In [17] it is proved that, if the rank of the derivative matrix, formed from differentiating $\kappa_i$ with respect to the parameters, is equal to the number of parameters, then the system is at least locally identifiable.

There are several other methods that are used to determine if a compartment model is identifiable or not. For example, [18] present a Markov parameter matrix approach involving a derivative matrix and a rank test. Another method for determining identifiability in compartment models is the Taylor series approach of [19], which is also applicable to non-linear compartment models, and for which the rank Jacobian test was introduced in [3]. There is also a ‘similarity of transform’ approach [6, 20], extended to non-linear compartment models in [21]. The use of a Jacobian matrix and a rank test for this similarity of transform approach was developed in [22].

**Example 2 Simple Linear Compartment Model.** We consider a simple compartment model, first considered in [16], and used subsequently by many other authors. In this compartment model,

$$\frac{dx_1}{dt} = -(\theta_1 + \theta_2)x_1 + \theta_3x_2 + u, \quad \frac{dx_2}{dt} = \theta_2x_1 - (\theta_3 + \theta_4)x_2, \quad \text{and} \quad y = x_1,$$

with $x(0) = 0$, $u(0) = 1$ and $u(t) = 0$ otherwise. The model parameters are $\theta = \left[ \theta_1 \quad \theta_2 \quad \theta_3 \quad \theta_4, \right]$. The transfer function is

$$Q(s) = \frac{s + \theta_3 + \theta_4}{s^2 + s(\theta_1 + \theta_2 + \theta_3 + \theta_4) + \theta_1\theta_3 + \theta_1\theta_4 + \theta_2\theta_4},$$

with non-constant coefficients

$$\kappa_1 = \theta_3 + \theta_4, \quad \kappa_2 = \theta_1 + \theta_2 + \theta_3 + \theta_4, \quad \kappa_3 = \theta_1\theta_3 + \theta_1\theta_4 + \theta_2\theta_4. \quad (2)$$
As there are three equations and four unknowns, obviously there are an infinite number of solutions to equations (2). We can check this result formally by forming the rank of appropriate derivative matrices. One derivative matrix can be formed by differentiating equations (2) with respect to the parameters. This derivative matrix is given in Table 2. An alternative derivative matrix can be formed from a Taylor series expansion of \( y(t) \), the first 3 non-constant coefficients of which are

\[
\begin{align*}
  y^{(2)}(0) &= -\theta_1 - \theta_2 \\
  y^{(3)}(0) &= (-\theta_1 - \theta_2)^2 + \theta_3 \theta_2 \\
  y^{(4)}(0) &= (-\theta_1 - \theta_2)\left\{(-\theta_1 - \theta_2)^2 + \theta_3 \theta_2\right\} + \theta_3 \theta_2 (-\theta_1 - \theta_2 - \theta_3 - \theta_4)
\end{align*}
\]

The Taylor series expansion is infinite. However, for a linear system of compartment models with \( n \) compartments, only the first \((2n - 1)\) derivatives, \( y^{(k)}(0) \) \((k \geq 2)\) are required [23, 24]. As in this example there are two compartments, we only need the \( y^{(k)}(0) \) given by equation (3). The resulting derivative matrix is also given in Table 2, and again has rank 3. In this instance the Markov parameter matrix approach, [18], results in the same derivative matrix as the Taylor series approach.

Mention of [23] prompts us to acknowledge substantial related work in the literature of dynamic systems; here we allude to just a sample of the more closely-related papers. [23] attributes the link between structural identifiability and the rank of the Taylor-series-related derivative matrix to pp.163–164 of [25]. [26] considers linear time-invariant systems like compartment models, with the square matrix \( A \) unknown and \( B \) and \( C \) known, as is the case with example 2. It is assumed that the eigenvalues of \( A \) are distinct and ‘known from experiment’, so identifiability of \( A \) is equivalent to that of the (non-singular) matrix of eigenvectors; an approach based on a Jacobian matrix is developed. For the same sort of model, [27] compares four methods, all using matrices whose rank properties are the same as those of Jacobian matrices but which ‘make the calculation of the determinant easier’. Again \( B \) and \( C \) are assumed
Table 2: Derivative matrices for the simple compartment model. \( \kappa \) is what is differentiated with respect to the parameters to form the derivative matrix. In both cases the derivative matrix is of rank 3

<table>
<thead>
<tr>
<th>( \kappa )</th>
<th>Derivative Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>eqn (2)</td>
<td>( D = \begin{bmatrix} 0 &amp; 1 &amp; \theta_3 + \theta_4 \ 0 &amp; 1 &amp; \theta_4 \ 1 &amp; 1 &amp; \theta_1 \ 1 &amp; 1 &amp; \theta_1 + \theta_2 \end{bmatrix} )</td>
</tr>
<tr>
<td>eqn (3)</td>
<td>( D = \begin{bmatrix} -1 &amp; 2\theta_1 + 2\theta_2 &amp; -3\theta_1^2 - 6\theta_1\theta_2 - 3\theta_2^2 - 2\theta_3\theta_4 \ -1 &amp; 2\theta_1 + 2\theta_2 + \theta_3 &amp; 3\theta_1^2 - 6\theta_1\theta_2 - 3\theta_2^2 - 4\theta_2\theta_3 - 2\theta_1\theta_3 - \theta_3^2 - \theta_3\theta_4 \ 0 &amp; \theta_2 &amp; -2\theta_1\theta_2 - 2\theta_2\theta_3 - \theta_2\theta_4 \ 0 &amp; 0 &amp; -\theta_2\theta_3 \end{bmatrix} )</td>
</tr>
</tbody>
</table>

known. Approach 1 is the Laplace transform approach of [16], Approach 2 is the Markov matrix approach of [18], which is noted to be equivalent to the Taylor series approach, Approach 3 is the approach in [26] and Approach 4 is that developed in [28], which we have already cited. It is shown in [27] that the approaches are in theory equivalent and investigates relationships between the determinants, which are zero in identical conditions but calculation of which can vary considerably in complexity.

The above paragraph restricts attention to linear systems. [29] consider continuous-time nonlinear dynamic models and define sensitivities as derivatives of states \( (x) \) or outputs \( (y) \) with respect to parameters; sensitivity matrices play the part of our derivative matrices. They then assume that the outputs follow a nonlinear regression model on \( t \), and relate identifiability to non-singularity of the information matrix. They calculate the sensitivity matrix of the set of derivatives of the output, of orders up to a particular level, and they relate investigation of its non-singularity to the Taylor series method for determining system identifiability. [30] consider solutions of sets of parametric nonlinear differential equations \( (\dot{x} = f(x, \theta, t)) \) and define the sensitivity matrix as the matrix of first derivatives of elements of \( x \) with respect to elements of \( \theta \).
sensitivity matrix consequently satisfies a system of ordinary differential equations, as it did in [29]. In some dynamic systems problems, the direct evaluation of the derivatives that represent sensitivities turns out to be computationally expensive, because of the temporal dependence implied by the model. However, the duality-based adjoint method can obviate this difficulty dramatically; see [31] and especially Section 5 of [32] for an introduction to the adjoint method.

In spite of what is a large literature, we found no discussion of reparameterisation or the other methodological aspects we consider in the rest of this paper.

1.3.2. Determining estimable parameters for parameter redundant models.

The rank of a derivative matrix gives more information than just whether or not a model is parameter redundant. The rank is equal to the number of parameters that are estimable and the derivative matrix can be used to determine exactly what is estimable, using results from [33]. If a model is parameter redundant then the rank of \( D \) is equal to the number of estimable parameters and the model is said to have deficiency \( d = p - \text{rank}(D) \). It is possible to tell which, if any, of the original parameters are estimable by solving \( \alpha(\theta)^T D(\theta) = 0 \). In this case there are \( d \) solutions to \( \alpha(\theta)^T D(\theta) = 0 \), labelled \( \alpha_j(\theta) \), for \( j = 1, \ldots, d \), with individual entries \( \alpha_{ij}(\theta) \). Any \( \alpha_{ij}(\theta) \) which are zero for all \( j \) correspond to a parameter which is estimable [33]. In order to find other parameter combinations which are also estimable, we need to solve the system of linear first-order partial differential equations \( \sum_{i=1}^{p} \alpha_{ij} \partial f/\partial \theta_i = 0, \ j = 1 \ldots r \) [33]. Also known as Lagrange equations, these are familiar from the analysis of linear stochastic models - see for example page 158 of [34]. A similar method for compartment models has also been developed in [22, 35].

**Example 1 - CJS model continued.** Recall that in this case the rank of the derivative matrix was 5 and that there were 6 parameters in the model. Therefore there are 5 estimable parameters and the model has deficiency 1. The single solution of \( \alpha^T D = 0 \) is \( \alpha^T = [0, 0, -\phi_3/p_4, 0, 0, 1] \). From the positions of
the zeros relative to the order of differentiation in the derivative matrix, we can see that \( \phi_1, \phi_2, p_2 \) and \( p_3 \) are estimable, but \( \phi_3 \) and \( p_4 \) are not. The remaining estimable term then results from solving the partial differential equation

\[
- \frac{\partial f}{\partial \phi_3} \frac{\partial f}{\partial p_4} = 0,
\]

the solution of which shows we can estimate \( \phi_3 p_4 \), as already observed. \( \square \)

For recent research in stochastic models for carcinogenesis see [36, 37].

2. Exhaustive summaries and their use

As shown in Section 1.3, models in different areas of application may have a variety of different starting points for determining the parametric structure of models. In Tables 1 and 2 different parameter vectors are differentiated to obtain the derivative matrices. What is differentiated is the key to determining parametric structure of models. In order to provide a general framework we call this quantity an exhaustive summary. Exhaustive summaries for compartment models are defined in [6, 20] and this can be extended to any parametric model. An exhaustive summary is a set of parameter combinations that uniquely defines the model, and a formal definition is given below, adapted from [6].

**Definition 4.** A parameter vector \( \kappa(\theta) \) is an exhaustive summary if knowledge of \( \kappa(\theta) \) uniquely determines \( M(\theta) \).

We have already seen several examples of exhaustive summaries. In example 1 three alternative exhaustive summaries are used. The first is \( \kappa_1(\theta) = \mu \), where \( \mu \) is given by equation (1), and the other two are \( \kappa_2(\theta) = [\phi_1 p_2, \phi_2 p_3 p_4, \phi_1 \phi_2 \phi_3 p_3 p_4, \phi_2 p_3, \phi_2 \phi_3 p_4, \phi_3 p_4]^T \) and \( \kappa_3(\theta) = \ln\{\kappa_2(\theta)\} \). In example 2 two different exhaustive summaries are used, given by equations (4) and (5), the first finite and the second infinite:

\[
\kappa_1(\theta) = [\theta_3 + \theta_4, \theta_1 + \theta_2 + \theta_3 + \theta_4, \theta_1 \theta_3 + \theta_1 \theta_4 + \theta_2 \theta_4]^T \tag{4}
\]

\[
\kappa_2(\theta) = [-\theta_1 + \theta_2, (\theta_1 + \theta_2)^2 + \theta_3 \theta_2, -(\theta_1 + \theta_2)^3 - \theta_3 \theta_2 (2\theta_1 + 2\theta_2 + \theta_3), \ldots]^T. \tag{5}
\]
Exhaustive summaries are useful because they specify a particular aspect of the model, $M(\theta)$, which can be used to determine model parametric structure.

**Theorem 1.** A model is globally (locally) identifiable if (there exists a region such that) $\kappa(\theta) = \kappa(\theta') \Rightarrow \theta = \theta'$.

The proof of Theorem 1 follows from Definitions 1 and 4. We show how an exhaustive summary can be used to determine parametric structure in Section 2.1 below.

Note that for linear and non-linear compartment models, when the Taylor series approach is used, the exhaustive summary is formed from every term of the Taylor series expansion. As pointed out in [6], this is an infinite exhaustive summary. In order to determine identifiability using symbolic algebra, only a finite number of $y^{(k)}$ terms need be considered, and in general this finite number of terms does not form an exhaustive summary. However, as observed above, upper bounds for the finite number of $y^{(k)}$ ($k \geq 2$) needed to determine model parametric structure are given in [24].

### 2.1. Detecting Parameter Redundancy using Exhaustive Summaries

In a similar way to [5, 22], and other references mentioned above, the basis of determining parametric structure is to form an appropriate derivative matrix, $D = [\partial \kappa_j / \partial \theta_i]$, where $\kappa_j$ is the $j$th element of the exhaustive summary, and $\theta_i$ is the $i$th of $p$ parameters. The derivative matrix can be used to examine parameter redundancy or identifiability as laid out in the following theorem.

**Theorem 2.** Testing parameter redundancy.

a. (i) If $D$ has rank equal to $p$ then the model is full rank.

(ii) If the rank of $D$ is equal to $q < p$, then the model is parameter redundant. There are $q$ estimable parameters and the model has deficiency $d = p - q$.

b. If the model is parameter redundant the estimable parameters can be determined by solving $\alpha(\theta)^T D(\theta) = 0$, which has $d$ solutions, labelled $\alpha_j(\theta)$ for $j = 1 \ldots d$, with individual entries $\alpha_{ij}(\theta)$. Any $\alpha_{ij}(\theta)$ which are zero
for all solutions correspond to a parameter, \( \theta_i \), which is estimable. The solutions of the system of linear first-order partial differential equations (PDEs),

\[
\sum_{i=1}^{p} \alpha_{ij} \frac{\partial f}{\partial \theta_i} = 0, \quad j = 1 \ldots r,
\]

form the set of estimable parameters. Parameterised in terms of the estimable parameters, the model is full rank.

Theorem 2 is a natural extension of similar theorems given in [5, 33]. Proofs follow the same lines as in those papers.

Example 1 CJS Model Continued. Theorem 2 has already been demonstrated for example 1 in Section 1.3, where it was shown that the CJS model is parameter redundant with deficiency 1 and that the set of estimable parameters is \( \{ \phi_1, \phi_2, p_2, p_3, \phi_3 p_4 \} \). The exhaustive summaries and their derivative matrices are given in Table 1. The simplest exhaustive summary to use is \( \kappa_3(\theta) = \ln(\kappa_2(\theta)) \).

Example 2 Simple Linear Compartment Model Continued. Part (b) of Theorem 2 has also already been demonstrated for example 2 in Section 1.3, where it was shown that the model was parameter redundant. The simplest exhaustive summary to use is that given by equation (4) and its derivative matrix is presented in Table 2. As the rank of the derivative matrix was 3 and there were 4 parameters in this model the deficiency is 1. We can find exactly what is estimable by solving \( \alpha^T D = 0 \). This gives \( \alpha^T = [-\theta_2/\theta_3, \theta_2/\theta_3, -1, 1] \). Solving the PDE

\[
-\frac{\partial f}{\partial \theta_1} \frac{\theta_2}{\theta_3} + \frac{\partial f}{\partial \theta_2} \frac{\theta_2}{\theta_3} - \frac{\partial f}{\partial \theta_3} + \frac{\partial f}{\partial \theta_4} = 0
\]

gives the estimable parameter combinations \( \theta_1 + \theta_2, \theta_2 \theta_3, \theta_3 + \theta_4 \). This can be readily appreciated from equation (1), where we can see that \( \kappa \) is a function of these three parameter combinations.
2.2. Extension Theorem

The method of Theorem 2 requires calculating the symbolic rank of the derivative matrix. If the derivative matrix is large and/or with complicated algebraic expressions, then it may not be possible to find the symbolic rank. One solution is to simplify the exhaustive summary, so that the derivative matrix is structurally simpler, and we consider this in Section 3. Another method of simplifying the calculation is to consider the smallest structural version of a model and then extend the size of that model whilst maintaining the model structure. This idea is advanced in [5] for full rank product-multinomial models and extended in [7] to parameter redundant product-multinomial models. This is generalised in Theorem 3 below.

Consider a model whose parametric structure is being examined using exhaustive summary $\kappa_1(\theta_1)$, with parameters $\theta_1$. The derivative matrix is $D_1(\theta_1) = \frac{\partial \kappa_1}{\partial \theta_1}$. This model is then extended, adding extra parameters, $\theta_2$, and the exhaustive summary is extended to be $\kappa(\theta') = [\kappa_1(\theta_1), \kappa_2(\theta')]$, with $\theta' = [\theta_1, \theta_2]$. The derivative matrix of the extended model is

$$D = \begin{bmatrix} D_1(\theta_1) & D_{2,1}(\theta_1) \\ 0 & D_{2,2}(\theta_2) \end{bmatrix}$$

with $D_{2,1} = \frac{\partial \kappa_2}{\partial \theta_1}$ and $D_{2,2} = \frac{\partial \kappa_2}{\partial \theta_2}$.

**Theorem 3.** If the original model is full rank (i.e $D_1$ is full rank) and $D_{2,2}$ is full rank, then the extended model is full rank also.

This approach can often be generalised to all models of the same type using induction. Proof of Theorem 3 follows from the fact that, as $D_1$ and $D_{2,2}$ are full rank, $D$ is also full rank, as shown in [5].

**Remark 1.** If the original model is not full rank, we first need to find a reparameterisation of the model that is full rank. Then Theorem 3 can be applied to the reparameterised model, so that deficiency of the general model can be deduced. Extension of parameter redundant models is considered explicitly in [7].
Example 3 - Recovery of Marked Animals. A model for the recovery of dead marked animals is defined by the matrix

\[ Q = \begin{bmatrix}
\bar{\phi}_{1,1} \lambda_1 & \phi_{1,1} \bar{\phi}_a \lambda_a & \phi_{1,1} \phi_a \bar{\phi}_a \lambda_a \\
0 & \bar{\phi}_{1,2} \lambda_1 & \phi_{1,2} \bar{\phi}_a \lambda_a
\end{bmatrix}, \]

where \( Q_{ij} \) is the probability of an animal being marked in year \( i \) and reported dead in year \( j \), for two years of marking and three years of recovery. The parameters are \( \theta = [\phi_{1,1}, \phi_{1,2}, \phi_a, \lambda_1, \lambda_a] \), where \( \phi_{1,1} \) is the probability that a first-year animal survives a year within the first year of the study, \( \phi_{1,2} \) is the probability that a first-year animal survives a year within the second year of the study, \( \phi_a \) is the probability that an adult animal survives a year for any year of the study, \( \lambda_1 \) is the reporting probability for animals in their first year of life and \( \lambda_a \) is the reporting probability for adult animals [38]. As this is a product-multinomial model, an exhaustive summary is the non-zero entries of \( Q \) [5]. The derivative matrix, (7), is of full rank 5:

\[ D_1 = \begin{bmatrix}
-\lambda_1 & \bar{\phi}_a \lambda_a & \phi_a \bar{\phi}_a \lambda_a & 0 & 0 \\
0 & 0 & 0 & -\lambda_1 & \bar{\phi}_a \lambda_a \\
0 & -\phi_{1,1} \lambda_a & \phi_{1,1} \bar{\phi}_1 \lambda_a - \phi_{1,1} \phi_1 \lambda_a & 0 & -\phi_{1,2} \lambda_a \\
\bar{\phi}_{1,1} & 0 & 0 & \bar{\phi}_{1,2} & 0 \\
0 & \phi_{1,1} \bar{\phi}_1 & \phi_{1,1} \phi_a \bar{\phi}_a & 0 & \phi_{1,2} \bar{\phi}_a
\end{bmatrix}. \tag{7} \]

First consider adding an extra year of marking animals. This results in the extra exhaustive summary entry \( s_+ = (1 - \phi_{1,3}) \lambda_1 \) and extra parameter \( \theta_2 = \phi_{1,3} \). As \( D_{2,2} = -\lambda_1 \), which is obviously full rank, the extended model is also full rank. Next consider adding an extra year of recovering animals. This results in the extra exhaustive summary terms \( s_+ = [\phi_{1,1} \phi_a \bar{\phi}_a \lambda_a, \phi_{1,2} \phi_a \bar{\phi}_a \lambda_a, \phi_{1,3} \phi_a \lambda_a]^T \). Whilst this adds extra entries to the exhaustive summary, there are no additional parameters so it is trivially full rank. This result can be extended to any number of years of marking and recovery by induction. □

Other applications of the extension theorem can be seen later, in the third treatment of example 1 in Section 3 and in examples 6 and 7 in Section 4.
2.3. Essentially Full Rank, Conditionally Full Rank and Near Parameter Redundant Models

It is possible to demonstrate that a model is not parameter redundant by showing that the derivative matrix, $D$, of an appropriate exhaustive summary is full rank. However we know from Definition 3 that it may not be full rank for all values of the parameters.

[39] provide an important advance by showing that it is possible to distinguish between particular essentially and conditionally full rank models by taking a decomposition of the derivative matrix. Here we extend this approach by providing a general decomposition method that can be used to determine whether a model is essentially or conditionally full rank. The decomposition that we employ is a (modified) PLUR decomposition.

**Theorem 4.** For a full-rank model, write $D = PLUR$, where $P$ is a permutation matrix, $L$ is a lower triangular matrix with ones on the diagonal, $U$ is an upper triangular matrix and $R$ is a matrix in reduced echelon form. The model is parameter redundant at $\theta$ if and only if $\text{Det}(U) = 0$ at a point $\theta \in \Omega$ and $R$ is defined at $\theta$.

Proof of Theorem 4 is given in Appendix A.1 and can also be deduced from [40].

**Remark 2.** If $R$ is not defined at $\theta$ then a further PLUR decomposition, of $D$ evaluated at $\theta$, is required.

It is also possible for a full rank model to be classified as parameter redundant in practice, and we define this as near parameter redundancy. Unlike earlier work, near parameter redundancy is a consequence of model-fitting.

**Definition 5.** A near parameter redundant model is one that is formally full rank, but might be classed as parameter redundant by a numerical method because it is close in some sense to a model that is parameter redundant.
Near parameter redundant models may result in unstable inference, see for example [38].

**Remark 3.** Theorem 4 can also be used to check for possible near parameter redundancy if the determinant of $U$ is near to zero, in some metric, for particular values of $\theta$. In this case a full rank model is near parameter redundant at $\theta$.

**Remark 4.** Theorem 4 and Remark 3 also apply as $\det(U)$ approaches infinity.

**Example 3 - Recovery of Marked Animals Continued.** A PLUR decomposition of the derivative matrix given by (7) results in $R$ being equal to the identity matrix (and hence always defined), and $\det(U) = \lambda_1 \phi_{1,1} \lambda_a^2 (1 - \phi_a)^3 (\phi_{1,1} - \phi_{1,2})$. This model is conditionally full rank. As well as being parameter redundant at certain boundary values for parameters ($\lambda_1 = 0, \phi_{1,1} = 0, \lambda_a = 0, \phi_a = 1$) which are unlikely to occur in practice, the model is also parameter redundant if $\phi_{1,1} = \phi_{1,2}$. This model is near parameter redundant in practice if $\phi_{1,1}$ is estimated as close to $\phi_{1,2}$, as demonstrated in [38]. □

In example 3 we may be interested in a model with only one first-year survival parameter, $\phi_1 = \phi_{1,1} = \phi_{1,2}$, and it is natural to consider whether or not this model is still full rank. For any sub-model formed from linearly constraining parameters, we have the result of Theorem 5.

**Theorem 5.** If a full rank model has any parameter redundant sub-models, formed from taking linear constraints on the elements of $\theta$, then they will appear as solutions of $\det(U) = 0$. The deficiency of the sub-model is $d_U$, where $d_U$ is the deficiency of $U$ evaluated with the sub-model constraints applied.

For proof of Theorem 5 see Appendix A.2.

**Example 3 - Recovery of Marked Animals Continued.** The sub-model with $\phi_{1,1} = \phi_{1,2} = \phi_1$ is known to be parameter redundant [41], but this parameter
redundancy is also picked up as a solution of $\text{Det}(U) = 0$. The rank of $U$ evaluated at $\phi_{1,1} = \phi_{1,2}$ is 4, resulting in a deficiency of 1. Therefore this sub-model has deficiency 1. However the model $\phi_{1,1} = \phi_a$, with a separate $\phi_{1,2}$, is still full rank as $\phi_{1,1} = \phi_a$ does not result in $\text{Det}(U) = 0$. □

We can combine the extension theorem (Theorem 3) with Theorem 4 to show that, if there are any points in $\Omega$ for which a model is not full rank, or sub-models of a full rank model that are parameter redundant, then in general that finding naturally extends to larger models of the same model structure. Consider again a model whose parametric structure is being examined using exhaustive summary $\kappa_1(\theta_1)$, with parameters $\theta_1$. The derivative matrix, $D_1 = [\partial \kappa_1/\partial \theta_1]$, is full rank and can be decomposed as previously, in the form $P_1L_1U_1R_1$, where $P_1$ is a permutation matrix, $L_1$ is a lower diagonal matrix with ones on the diagonal, $U_1$ is an upper triangular matrix and $R_1$ is a matrix in reduced echelon form. This model is then extended, adding extra parameters, $\theta_2$, and the extra exhaustive summary terms $\kappa_2$. If we write the derivative matrix $D_2 = [\partial \kappa_2/\partial \theta_2] = P_2L_2U_2R_2$, then we have the following result for the new model.

**Theorem 6.** If $\text{Det}(U_1) = 0$ or $\text{Det}(U_2) = 0$ at a point $\theta$ in the parameter space then the new extended model is parameter redundant at $\theta$.

Proof of Theorem 6 follows from Theorems 3 and 4, and the fact that we can write the derivative matrix of the new model in the form

$$D_n = P_nL_nU_nR_n = \begin{bmatrix} P_1L_1U_1R_1 & X \\ 0 & P_2L_2U_2R_2 \end{bmatrix},$$

where $X$ denotes terms that are not relevant to the argument. Therefore $\text{Det}(U_n) = \text{Det}(U_1) \times \text{Det}(U_2) = 0$.

**Example 3 - Recovery of Marked Animals Continued.** Previously we have shown that $\text{Det}(U) = \lambda_1\phi_{1,1}\lambda_2^3(1 - \phi_a)^3(\phi_{1,1} - \phi_{1,2})$ and that adding an extra year of marking animals results in the extra exhaustive summary entry $s_+ = 19$.
(1 − ϕ1,3)λ1 and an extra parameter θ2 = ϕ1,3 and that adding an extra year of recovering animals results in no extra parameter. For this extension Det(U2) = −λ1, so that the new model has Det(Un) = −λ1,1λ2,1λ3,1(1 − ϕa)3(ϕ1,1 − ϕ1,2).

Therefore, in general for n ≥ 2 years of ringing and more than 2 years of recovery, Det(U) = (−1)n−1λ1,1n−1λ2,1n−1(1 − ϕa)3(n − 1n−1)(ϕ1,1 − ϕ1,2).

This shows that the model will be parameter redundant if ϕ1,1 = ϕ1,2, but that ϕ1,3, ϕ1,4,... can take any values and this will not affect whether or not the model is parameter redundant.

□

Theorem 7.

(i) If a model is essentially full rank it is at least locally identifiable.

(ii) If a model is conditionally full rank, it is at least locally identifiable almost everywhere. If we exclude the points found in Theorem 4, the model is at least locally identifiable.

Proof follows from Theorem 4 of [5].

Remark 5. A model that is conditionally full rank has points in Ω where the model is parameter redundant. These points correspond to sub-models that need to be considered separately. If the sub-model results from a linear constraint on the parameters, then Theorem 5 can be applied. If the model results from setting θi = 0, then the sub-model needs to be considered separately, and could still be locally identifiable. See for example [42].

3. Reduced-form exhaustive summary

As has already been observed, some models are too structurally complex for the symbolic method of testing for parameter redundancy to apply. For example, [43, 44] applied the symbolic method to their models and discovered that Maple ran out of memory trying to calculate the rank of the derivative matrices that they adopted. Trying to choose a simple exhaustive summary and/or applying the extension theorem are approaches that can be used in such situations, as
described above. If this approach is not sufficient, then a new simpler exhaustive summary may be derived from any reparameterisation that simplifies the model structure. We show below how we can derive and apply reduced-form exhaustive summaries, by conveniently making use of earlier results in the paper.

A useful exhaustive summary is termed a reduced-form exhaustive summary.

Definition 6. A reduced-form exhaustive summary is a reparameterisation \( r(\theta) \) such that the derivative matrix \( D_r = \left[ \frac{\partial \kappa_j(r)}{\partial r_i} \right] \) is full rank.

For any model, the reduced-form exhaustive summary is not unique. Let \( \kappa(\theta) \) be an exhaustive summary and choose a reparameterisation \( s(\theta) \). Typically the reparameterisation would be independent in terms of \( \theta \), so that \( \text{rank}(\partial s/\partial \theta) = \text{dim}(s) \). Rewriting \( \kappa(\theta) \) in terms of \( s \) gives \( \kappa(s) \) and the intention is that \( \kappa(s) \) is structurally simpler than \( \kappa(\theta) \). The derivative matrix is now \( D_s = \left[ \frac{\partial \kappa_j}{\partial s_i} \right] \). If rank \( (\partial s/\partial \theta) = \text{dim}(s) \), the derivative matrix \( D_s \) gives us sufficient information to determine the parameter redundancy of the original model of interest, but can be computationally much simpler than the original derivative matrix \( D \). Further, regardless of rank \( (\partial s/\partial \theta) \), \( s \) can be used to find a reduced-form exhaustive summary. These results are laid out in Theorem 8.

Theorem 8. Reparameterisation.

Let \( \text{rank}(D_s) = r_s \), and \( p_s = \text{dim}(s) \). Then the following hold.

a. (i) If \( r_s = p_s \), \( s \) is a reduced-form exhaustive summary. The results of Theorems 2 to 5 can then be applied to \( s \), to examine model parametric structure.

(ii) If \( r_s < p_s \) then \( s \) is not a reduced-form exhaustive summary. A reduced-form exhaustive summary may be found by first solving \( \alpha^T D_s = 0 \) and then solving the appropriate partial differential equations as in Theorem 2(b).

b. If \( \text{rank}(\partial s/\partial \theta) = p_s \), the number of estimable parameters is equal to \( r_s \).

If \( r_s = p \) then the model in terms of \( \theta \) is full rank. If \( r_s < p \) the model in terms of \( \theta \) is parameter redundant.

The proof of Theorem 8 is given in Appendix A.3.
Remark 6. If \( \text{rank}(\partial s/\partial \theta) \neq p_s \), an exhaustive summary may be formed by solving \( \alpha^T D_s = 0 \) and then solving the appropriate PDEs. However, part (b) of Theorem 8 does not necessarily apply; instead the new exhaustive summary should be used with Theorem 2 to determine whether or not the model is parameter redundant. An illustration of this is provided in [8], where the reparameterisation of \( s \) introduces an additional redundancy, in order to simplify model structure and allow the use of computer algebra.

Applying Theorem 8 is effectively a two-stage procedure. The first stage is to choose a (finite) reparameterisation, \( s \), that simplifies a model. We rewrite the original exhaustive summary, \( \kappa(\theta) \), in terms of the chosen reparameterisation, \( s \), and find its derivative matrix \( D_s \). If \( \text{rank}(\partial s/\partial \theta) = \text{dim}(s) \), as is often the case, the rank of \( D_s \) tells us how many estimable parameters there are in the model. The procedure can stop here if the interest only lies in whether or not the model is parameter redundant. Alternatively we can move on to the second stage, which involves solving the appropriate set of PDEs to find a reduced-form exhaustive summary, \( r \) (if \( s \) is not already a reduced-form exhaustive summary). Then \( D_2 = [\partial r_j/\partial \theta_i] \) can be used to find the estimable set of parameters if the model is parameter redundant, or if the model is full rank Theorem 4 can be applied to either \( D_2 \) or \( D_s \) to determine if the model is conditionally full rank, etc.

To demonstrate how reparameterisation can be used, we first provide a possible reparameterisation of example 1. Although reparameterisation is not needed for the CJS model, this simple model allows us to illustrate the approach and finding a new reduced-form exhaustive summary allows for the application of the extension theorem. A further example is then given where the reparameterisation method is necessary in order to determine the number of estimable parameters in a more complex model.

Example 1 - CJS Model Continued. A possible reparameterisation of the
CJS model is given below, and is equivalent to that in the last example:

\[
\mathbf{s} = [s_1, s_2, s_3, s_4, s_5]^T = [\phi_1 p_2, \phi_1 \bar{p}_2, \phi_2 p_3, \phi_2 \bar{p}_3, \phi_3 p_4]^T.
\]

(8)

In this case, \(\text{rank}(\partial \mathbf{s}/\partial \mathbf{\theta}) = p_s\). The exhaustive summary \(\mathbf{\kappa}_2(\mathbf{\theta}) = \ln\{\mathbf{\kappa}_2(\mathbf{\theta})\}\) can be rewritten in terms of the reparameterisation \(\mathbf{s}\), giving

\[
\mathbf{\kappa}(\mathbf{s}) = [\ln(s_1), \ln(s_2 s_3), \ln(s_2 s_4 s_5), \ln(s_3), \ln(s_4 s_5), \ln(s_5)]^T.
\]

The derivative matrix \(\mathbf{D}_s\) is then

\[
\mathbf{D}_s = \left[\frac{\partial \mathbf{\kappa}_j(\mathbf{s})}{\partial s_i}\right] = \begin{bmatrix}
    s_1^{-1} & 0 & 0 & 0 & 0 & 0 \\
    0 & s_2^{-1} & s_2^{-1} & 0 & 0 & 0 \\
    0 & s_3^{-1} & 0 & s_3^{-1} & 0 & 0 \\
    0 & 0 & s_4^{-1} & 0 & s_4^{-1} & 0 \\
    0 & 0 & s_5^{-1} & 0 & s_5^{-1} & s_5^{-1} \\
\end{bmatrix},
\]

which has full rank 5. We note that this form is necessarily simpler than any provided in Table 1. Therefore, by Theorem 8b we have again shown that this model is parameter redundant with deficiency 1. Further, the reparameterisation given by equation (8) is a reduced form exhaustive summary.

In addition we can use the extension theorem to show that a general CJS model with \(r = c = n\) has a reduced-form exhaustive summary,

\[
\mathbf{s}^e = [\phi_1 p_2, \phi_1 \bar{p}_2, \ldots, \phi_{n-1} p_n, \phi_{n-1} \bar{p}_n, \phi_n p_{n+1}]^T.
\]

(9)

Consider first extending the CJS model from \(r = 3\) and \(c = 3\) to \(r = 4\) and \(c = 4\). The extra reparameterisation terms are \(\mathbf{s}_+ = [s_6, s_7]^T = [\phi_3 \bar{p}_4 p_5]^T\) and the extra terms in the original exhaustive summary are \(\mathbf{k}_+(\mathbf{s}_+) = [\ln(s_2 s_4 s_6 s_7), \ln(s_4 s_6 s_7), \ln(s_6 s_7), \ln(s_7)]^T\). The derivative matrix,

\[
\mathbf{D}_{s,+} = \left[\frac{\partial \mathbf{\kappa}_{+,j}(\mathbf{s})}{\partial s_{+,i}}\right] = \begin{bmatrix}
    s_6^{-1} & s_6^{-1} & s_6^{-1} & 0 \\
    s_7^{-1} & s_7^{-1} & s_7^{-1} & s_7^{-1} \\
\end{bmatrix},
\]

has full rank 2, and therefore by the extension theorem (Theorem 3) the extended model is also full rank, implying that the extended reparameterisation
is also a reduced-form exhaustive summary, and by induction the CJS model with \( r = c = n \) has a reduced-form exhaustive summary given by equation (9). From reduced-form exhaustive summary (9) it then follows that the CJS model is always parameter redundant with deficiency 1.

Example 4 - Multi-state Capture-Recapture Model. [43] examine the parameter redundancy of a number of multi-state capture-recapture models for seabirds. They found the method of [5] intractable, and adopted an alternative approach. Their 4-state model for breeding success and failure has one state for success, one state for failure and one state for each of post-success and post-failure of breeding (which are both unobservable) and three different types of parameter for state \( i \): \( \sigma_i \), the probability of survival, \( \beta_i \), the probability of breeding given survival, and \( \gamma_i \), the probability of success given breeding. For this example the model is the log-likelihood given by

\[
\log L = \sum_{r=1}^{N-1} \sum_{c=r+1}^N \sum_{i=1}^4 \sum_{j=1}^4 m_{ij}^{(r,c)} \log \Psi_{ij}^{(r,c)} \equiv M(\theta),
\]

where \( N \) is the number of years of both marking and observation, \( m_{ij}^{(r,c)} \) is the number of individuals released in stage \( i \) at occasion \( r \) that are next captured in stage \( j \) at occasion \( c \) and where \( \Psi_{ij}^{(r,c)} \) is the probability that an individual released in stage \( i \) at time \( r \) is next captured in stage \( j \) at time \( c \), so that

\[
\Psi_{ij}^{(r,c)} = \begin{cases} 
(\Pi_{r+1} \Phi_r)^T & c = r + 1 \\
(\Pi_r \Phi_{c-1} (I - \Pi_{c-1}) \Phi_{c-2} \ldots (I - \Pi_{r+1}) \Phi_r)^T & c > r + 1 
\end{cases},
\]

with \( \Psi_{ij}^{(r,c)} = 0 \) for \( j = 3, 4 \) (as these states are unobservable), where the transition and recapture matrices are (respectively for all \( t \))

\[
\Phi_t = \begin{bmatrix}
\sigma_1 \beta_1 \gamma_1 & \sigma_2 \beta_2 \gamma_2 & \sigma_3 \beta_3 \gamma_3 & \sigma_4 \beta_4 \gamma_4 \\
\sigma_1 \bar{\beta}_1 \gamma_1 & \sigma_2 \beta_2 \bar{\gamma}_2 & \sigma_3 \beta_3 \bar{\gamma}_3 & \sigma_4 \beta_4 \bar{\gamma}_4 \\
\sigma_1 \bar{\beta}_1 & 0 & \sigma_3 \bar{\beta}_3 & 0 \\
0 & \sigma_2 \bar{\beta}_2 & 0 & \sigma_4 \bar{\beta}_4
\end{bmatrix}, \quad \Pi_t = \begin{bmatrix}
p_1 & 0 & 0 & 0 \\
0 & p_2 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}.
\]

For illustration, for the case where \( N = 5 \), the model has \( p = 14 \) parameters,

\[
\theta = [\sigma_1, \ldots, \sigma_4, \beta_1, \ldots, \beta_4, \gamma_1, \ldots, \gamma_4, p_1, p_2].
\]
The exhaustive summary, $\kappa$, used by [43] consists of the 40 non-zero elements of $\Psi$ with

$$
\kappa(\theta) = \begin{bmatrix}
p_1\sigma_1\beta_1\gamma_1 \\
p_2\sigma_1\beta_1(1 - \gamma_1) \\
p_1\sigma_2\beta_2\gamma_2 \\
p_2\sigma_2\beta_2(1 - \gamma_2) \\
p_1\sigma_1^2\beta_1^2\gamma_1^2(1 - p_1) + p_1\sigma_2\beta_2\gamma_2(1 - p_2)\sigma_1\beta_1(1 - \gamma_1) + p_1\sigma_3\beta_3\gamma_3\sigma_1(1 - \beta_1) \\
\vdots
\end{bmatrix}.
$$

It is clear that $\kappa(\theta)$ is an exhaustive summary as its specification determines $M(\theta)$. However, Maple cannot calculate the rank of the derivative matrix, $D = [\partial\kappa_j/\partial\theta_i]$. In order to find a reparameterisation to simplify the model, we observe that the parameters enter the model through the matrices $\Phi_t$ and $\Pi_t$, and we use their non-zero elements to form the reparameterisation $s$, giving

$$
s = \begin{bmatrix}
s_1 \\
s_2 \\
s_3 \\
s_4 \\
s_5 \\
s_6 \\
s_7 \\
s_8 \\
s_9 \\
s_{10} \\
s_{11} \\
s_{12} \\
s_{13} \\
s_{14}
\end{bmatrix} = \begin{bmatrix}
\sigma_1\beta_1\gamma_1 \\
\sigma_2\beta_2\gamma_2 \\
\sigma_3\beta_3\gamma_3 \\
\sigma_4\beta_4\gamma_4 \\
\sigma_1\beta_1(1 - \gamma_1) \\
\sigma_2\beta_2(1 - \gamma_2) \\
\sigma_3\beta_3(1 - \gamma_3) \\
\sigma_4\beta_4(1 - \gamma_4) \\
\sigma_1(1 - \beta_1) \\
\sigma_2(1 - \beta_2) \\
\sigma_3(1 - \beta_3) \\
\sigma_4(1 - \beta_4) \\
p_1 \\
p_2
\end{bmatrix}.
$$

Once again, $\text{rank}(\partial s/\partial \theta) = p_s$. The next stage is to rewrite $\kappa$ as a function of
\[ s, \text{ giving} \]
\[
\kappa(s) = \begin{bmatrix}
    s_1s_{13} \\
    s_5s_{14} \\
    s_2s_{13} \\
    s_6s_{14} \\
    (s_1^2 - s_7^2s_{13} + s_2s_{5}s_{14} + s_3s_9)s_{13} \\
    \cdots
\end{bmatrix}.
\]

We can see that \( \kappa(s) \) is structurally simpler than \( \kappa(\theta) \). The derivative matrix \( D_s \) is given by
\[
D_s = \left[ \frac{\partial \kappa_j(s)}{\partial s_i} \right] = \begin{bmatrix}
    s_{13} & 0 & 0 & 0 & (2s_1 - 2s_1s_{13})s_{13} & \cdots \\
    0 & 0 & s_{13} & 0 & (s_5 - s_5s_{14})s_{13} & \\
    0 & 0 & 0 & 0 & s_9s_{13} & \\
    0 & 0 & 0 & 0 & 0 & \\
    0 & s_{14} & 0 & 0 & (s_2 - s_2s_{14})s_{13} & \\
    \vdots
\end{bmatrix}
\]

and is found to have symbolic rank 12. Therefore, from Theorem 8b, the original model has 12 estimable parameters, and so is parameter redundant with deficiency 2. The reparameterisation above is not in reduced-form, but a reduced-form exhaustive summary can be found. Solving \( \alpha^T D_s = 0 \) gives \( \alpha_1 = [0, 0, 0, -s_4/s_{10}, 0, 0, 0, -s_8/s_{10}, 0, 1, 0, 0, 0, 0] \) and \( \alpha_2 = [0, 0, 0, -s_4/s_{10}, 0, 0, 0, -s_8/s_{10}, 0, 1, 0, 0, 0, 0] \). We can then solve the PDEs
\[
-\frac{\partial f}{\partial s_4} s_4 + \frac{\partial f}{\partial s_8} s_8 + \frac{\partial f}{\partial s_{10}} = 0
\]
\[
-\frac{\partial f}{\partial s_3} s_3 - \frac{\partial f}{\partial s_7} s_7 + \frac{\partial f}{\partial s_9} = 0
\]
to give a reduced-form exhaustive summary of

\[
\begin{bmatrix}
  s_1 \\
  s_2 \\
  s_5 \\
  s_6 \\
  s_{11} \\
  s_{12} \\
  s_{13} \\
  s_{14} \\
  s_7/s_3 \\
  s_8/s_4 \\
  s_3 s_9 \\
  s_4 s_{10}
\end{bmatrix} =
\begin{bmatrix}
  \sigma_1 \beta_1 \gamma_1 \\
  \sigma_2 \beta_2 \gamma_2 \\
  \sigma_1 \beta_1 (1 - \gamma_1) \\
  \sigma_2 \beta_2 (1 - \gamma_2) \\
  \sigma_3 (1 - \beta_3) \\
  \sigma_4 (1 - \beta_4) \\
  p_1 \\
  p_2 \\
  (1 - \gamma_3)/\gamma_3 \\
  (1 - \gamma_4)/\gamma_4 \\
  \sigma_3 \beta_3 \gamma_3 \sigma_1 (1 - \beta_1) \\
  \sigma_4 \beta_4 \gamma_4 \sigma_2 (1 - \beta_2)
\end{bmatrix}.
\]  

(10)

The extension theorem (Theorem 3) can be applied to show that this is a reduced-form exhaustive summary for any \( N \geq 5 \).

The reduced-form exhaustive summary given by equation (10) is evidently a much simpler exhaustive summary than the original exhaustive summary. This simpler reduced-form exhaustive summary can be used to find the estimable set of parameters, by solving \( \alpha^T D_2 = 0 \), where \( D_2 = [\partial r_j/\partial \theta_i] \), and the appropriate set of PDEs (as shown in Example4.mw). The estimable set of parameters is shown to be \( \sigma_1 \beta_1, \sigma_2 \beta_2, (1 - \beta_3)/\{\sigma_1 \beta_3 (1 - \beta_1)\}, \sigma_1 \sigma_3 \beta_3 (1 - \beta_1), (1 - \beta_4)/\{\sigma_2 \beta_4 (1 - \beta_2)\}, \sigma_2 \sigma_4 \beta_4 (\beta_2 - 1), \gamma_1, \gamma_2, \gamma_3, \gamma_4, p_1, p_2 \).

The reduced-form exhaustive summary given by equation (10) can also be used to determine the parametric structure of any 4-state model for this application. [43] consider constrained models, where parameters of the same type are set equal; for example, \( \beta_1 = \beta_2 \) and \( \beta_3 = \beta_4 \). Determining possible parameter redundancy under such constraints just involves forming the derivative matrix by differentiating the reduced-form exhaustive summary (10) evaluated at \( \beta_1 = \beta_2 \) and \( \beta_3 = \beta_4 \) with respect to the parameters, \( \theta = [\sigma_1, \ldots, \sigma_4, \beta_2, \beta_4, \gamma_1, \ldots, \gamma_4, p_1, p_2] \). In this case the derivative matrix has rank

27
11, so that this 12-parameter model is parameter redundant with deficiency 1 (the calculation for this and all of the constrained models considered in [43] are given in Example4.mw).

Reparameterisation can be used to seek a simple exhaustive summary. The trick in applying this method lies in the choice of the reparameterisation, \( s \). The reparameterisation needs to be chosen so that \( D_s = \left[ \frac{\partial \kappa_j(s)}{\partial s_i} \right] \) is simpler in structure than \( D = \left[ \frac{\partial \kappa_j(\theta)}{\partial \theta_i} \right] \). Typically a reparameterisation can be chosen by observing how parameters enter a model. In example 4 this was through the matrices \( \Phi_t \) and \( \Pi_t \), so their elements form a natural possible reparameterisation.

**Remark 7.** Theorem 8 requires the reparameterisation \( s \) to be finite. In the case of infinite exhaustive summaries, as may arise from a Taylor series expansion, a reparameterisation needs to be chosen such that \( s \) is finite, for Theorem 8b to apply.

For example, if only the first \( n \) terms of a Taylor series are considered, and a reparameterisation is chosen such that adding an extra term of the Taylor series adds one or more extra elements to \( s \), then Theorem 8b does not hold. This is because the first \( n \) terms of the Taylor series do not actually form an exhaustive summary, so that the true reparameterisation would be infinite. There are two possible solutions to this problem: the parameter redundancy can be determined from \( D_2 = \left[ \frac{\partial r_j}{\partial \theta_i} \right] \), where \( r \) is a reduced-form exhaustive summary formed from \( s \), or a different reparameterisation can be chosen such that adding an extra term of the Taylor series adds no extra element to \( s \). An example of this is shown in example 6 of Section 4.2 and in example 2 below.

**Example 2 - Simple Linear Compartment Model Continued.** This is another simple example that is used purely for illustration. We can simplify the Taylor series exhaustive summary, equation (5), by using the reparameterisation

\[
s = [s_1, s_2, s_3, s_4]^T = [-(\theta_1 + \theta_2), \theta_2, \theta_3, -(\theta_3 + \theta_4),]^T.
\]

28
This is a finite reparameterisation, yet it is possible to write any entry in the exhaustive summary given by equation (5) in terms of $s$.

Writing equation (5) in terms of $s$ gives

$$\kappa_2(s) = \begin{bmatrix} s_1 \\ s_1^2 + s_2s_3 \\ s_1(s_1^2 + s_2s_3) + s_3(s_1s_2 + s_2s_4) \\ \vdots \end{bmatrix},$$

and the derivative matrix is

$$D_s = \begin{bmatrix} 1 & 2s_1 & 3s_1^2 + 2s_2s_3 \\ 0 & s_3 & 2s_1s_3 + s_3s_4 \\ 0 & s_2 & 2s_1s_2 + s_2s_4 \\ 0 & 0 & s_2s_3 \end{bmatrix}.$$ 

The rank of the derivative matrix is 3. Therefore, $s$ is not a reduced-form exhaustive summary but a reduced-form exhaustive summary can be found, namely

$$r = [s_1, s_2s_3, s_4]^T = [- (\theta_1 + \theta_2), \theta_2 \theta_3, -(\theta_3 + \theta_4)],$$

which is the set of estimable parameters (multiplied by $-1$) found earlier. □

3.1. Local and Global Identifiability

Previously we have been primarily interested in determining whether a model is parameter redundant or full rank. If a model is essentially full rank, then it is at least locally identifiable. It is certainly possible that a full rank model can be globally identifiable, but this is more difficult to prove. As discussed in Section 1.3, a linear compartment model is globally identifiable if the solution to the moment invariant equations in terms of the parameters is unique. The moment invariant equations result from an exhaustive summary, and this approach can be extended to any exhaustive summary.
Theorem 9. A full rank model with a reduced-form exhaustive summary $r$ is globally identifiable if and only if there is a unique solution to the set of equations $k = r(\theta)$ in terms of $\theta$, where $\dim(k) = \dim(r)$.

Theorem 9 is an extension of the approach employed in the Laplace transform method used to detect identifiability in compartment models, see [17] page 2. A similar concept is discussed in [1] for linear reduced-form exhaustive summaries. Theorem 9 is a direct result of Theorem 1: if there is a unique solution to the set of equations $k = r(\theta)$, this means that $r(\theta) = r(\theta')$ implies that $\theta = \theta'$.

Example 3 - Recovery of Marked Animals Continued. The exhaustive summary is in reduced-form and the model is conditionally full rank (it is not full rank if $\phi_{11} = \phi_{12}$). The set of equations $k = s^{re}(\theta)$ is given by

\begin{align*}
k_1 & = (1 - \phi_{11})\lambda_1, & k_2 & = \phi_{11}(1 - \phi_a)\lambda_a, & k_3 & = \phi_{11}\phi_a(1 - \phi_a)\lambda_a, \\
k_4 & = (1 - \phi_{12})\lambda_1, & k_5 & = \phi_{12}(1 - \phi_a)\lambda_a,
\end{align*}

and has a unique solution,

\begin{align*}
\phi_{11} & = \frac{k_2(k_4 - k_1)}{k_5k_1 - k_4k_2}, & \phi_{12} & = \frac{k_5(k_4 - k_1)}{k_5k_1 - k_4k_2}, & \phi_a & = \frac{k_3}{k_2}, \\
\lambda_1 & = \frac{k_5k_1 - k_4k_2}{k_5 - k_2}, & \lambda_a & = \frac{k_1k_2 - k_4k_2 - k_3k_1 + k_3k_4}{k_1k_2 - k_4k_2 - k_3k_1 + k_3k_4},
\end{align*}

except if $\phi_{11} = \phi_{12}$. Therefore, this model is globally identifiable almost everywhere (the exception being if $\phi_{11} = \phi_{12}$, when the model is parameter redundant). Note that for maximum-likelihood estimation in this case, the model is saturated, so that the parameter expressions above are maximum-likelihood estimates when the $k_i$ are appropriate frequencies. □

Example 1 - CJS Model Continued. The reduced-form exhaustive summary of equation (11) leads directly to an explicit solution for the model parameters, when we treat the product of parameters $\phi_3p_4$ as a single parameter, in terms of $k_i$. Thus, in terms of this parameterisation, the CJS model is globally identifiable. In contrast to the last example, the model is not saturated in this case. □
Global identifiability in compartment models is assessed in [45] by examining whether or not the equations $k = r(\theta)$ have a unique solution by using methods of simplification, such as the Gröbner basis. Reduced-form exhaustive summaries can provide an alternative simplifying approach, as shown below in example 5.

4. Further Examples

In this section we look at a variety of further examples to illustrate the use of reparameterisation and exhaustive summaries as tools to simplify structurally complex models. In each case we draw conclusions using the theory presented above, thereby demonstrating the wide applicability of the approach of this paper.

4.1. Example 5 - global identifiability in a linear compartment model

A linear compartment model which has

\[
A = \begin{bmatrix}
-(k_{21} + k_{41}) & k_{12} & 0 & k_{14} \\
k_{21} & -(k_{12} + k_{32}) & k_{23} & 0 \\
0 & k_{32} & -(k_{23} + k_{43} + k_{03}) & k_{34} \\
k_{41} & 0 & k_{43} & -(k_{14} + k_{34} + k_{04})
\end{bmatrix},
\]

\[
B = \begin{bmatrix}
1 & 0 & 0 & 0
\end{bmatrix}^T,
\]

\[
C = \begin{bmatrix}
1/V_1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0
\end{bmatrix},
\]

with 11 parameters, $\theta = [k_{21}, k_{12}, k_{32}, k_{23}, k_{43}, k_{41}, k_{14}, k_{03}, k_{04}, V_1]$, is considered in [45], where it is shown to be globally identifiable. Here we use reduced-form exhaustive summaries as an alternative way of showing that this model is globally identifiable. In this example the exhaustive summary is the set of coefficients of the transfer function as described in Section 1.3 and example 2.
These coefficients are
\[
\kappa = \begin{bmatrix}
k_{14}k_{43}k_{32} + k_{14}k_{23}k_{12} + \ldots + k_{3}k_{4}k_{12} \\
k_{34}k_{32} + k_{32}k_{14} + k_{34}k_{12} + \ldots + k_{4}k_{32} \\
\vdots \\
k_{21}k_{32} + k_{34}k_{41}
\end{bmatrix};
\]
the exhaustive summary is given in full in [45]. However, it is not possible to calculate the symbolic rank of the derivative matrix \(D = \left[ \frac{\partial \kappa_j(\theta)}{\partial \theta_i} \right] \). A possible reparameterisation is
\[
s = \begin{bmatrix}
s_1 \\
s_2 \\
s_3 \\
s_4 \\
s_5 \\
s_6 \\
s_7 \\
s_8 \\
s_9 \\
s_{10} \\
s_{11}
\end{bmatrix} = \begin{bmatrix}
-(k_{21} + k_{41}) \\
-(k_{12} + k_{32}) \\
-(k_{21} + k_{43} + k_{33}) \\
-(k_{14} + k_{34} + k_{04}) \\
k_{12}k_{21} \\
k_{14}k_{11} \\
k_{23}k_{32} \\
k_{34}k_{43} \\
k_{21} \\
k_{41}k_{34} \\
V_1
\end{bmatrix}.
\]
This is based on [45], with the terms \([k_{34}k_{23}k_{14}, k_{32}k_{43}k_{14}, k_{21}k_{32}], \) also presented there, excluded, as these terms can be rewritten in terms of \(s\); it is derived using the topological properties of the graph associated with the model.
Writing the exhaustive summary in terms of \(s\) gives
\[
\kappa(s) = \begin{bmatrix}
s_2s_8 + s_4s_7 - s_2s_3s_4 \\
s_4s_3 + s_4s_2 - s_7 + s_3s_2 - s_8 \\
\vdots \\
s_{10} - s_5 - s_9s_2
\end{bmatrix}.
\]
The rank of \(D_s = \left[ \frac{\partial \kappa(s)}{\partial s_i} \right] \) is 11, so that \(s\) is a reduced-form exhaustive summary. As the original model has 11 parameters, this model is not parameter
redundant. If we solve the set of equations $s_i(\theta) = b_i$, for $i = 1, ..., 11$, we find that there is a unique solution for each of the parameters with $k_{21} = b_9$, $k_{12} = b_5/b_9$, $k_{32} = -(b_9b_2 + b_5)/b_9, ..., V_1 = b_{11}$. Hence the model is globally identifiable.

4.2. Example 6 - Bio-kinetic models of sludge respiration

The identifiability of models for the activated sludge process using a non-linear compartment model is examined in [46]. The Maple worksheet for this example is Example6.mw. [46] models the exogenous oxygen uptake, $U$ as

$$U = -\sum_{i=1}^{k} (1 - Y_i) \frac{dS_i(t)}{dt} \quad \text{with} \quad \frac{dS_i(t)}{dt} = -\frac{\mu_{max}iX}{Y_i} \frac{S_i(t)}{K_{mi} + S_i(t)}, \quad (11)$$

where $S_i$ is the $i$th of $k$ pollutants, and $Y_i$ is the fraction of the pollutant $S_i$ which is not oxidised. The first model we consider here is the single monod model with $k = 1$ and parameters $\theta = [Y_1, S_1(0), \mu_{max1}, K_{m1}, X]$. The second model we consider is the double monod model with $k = 2$ and parameters $\theta = [Y_1, S_1(0), \mu_{max1}, K_{m1}, Y_2, S_2(0), \mu_{max2}, K_{m2}, X]$.

Due to the combination $\mu_{max1}X/Y_1$, it is obvious that this model must be parameter redundant, but it is not as simple as determining that this is an identifiable parameter combination because the $Y_i$ also appears in $U$. [46] shows using the Taylor series expansion method that the single monod model is parameter redundant with deficiency 2. However, the complexity of the double monod model means that the computer runs out of memory when one tries to calculate the rank of the derivative matrix. Instead [46] examines the identifiability of the double monod model by transforming the nonlinear model to a linear model. Here we shall show how the Taylor series expansion method can still be used by applying Theorem 8.

First consider the simpler single monod model. The obvious exhaustive summary is the set of coefficients of a Taylor series expansion of equation (11)
with

$$\kappa(\theta) = \begin{bmatrix}
\frac{(1-Y_1)\mu_{max_1}XS_1(0)}{Y_1(K_{m_1}+S_1(0))} \\
\frac{-(1-Y_1)\mu_{max_1}X^2S_1(0)K_{m_1}}{Y_1(K_{m_1}+S_1(0))^3} \\
\frac{(1-Y_1)\mu_{max_2}X^3S_1(0)K_{m_1}(K_{m_1}-2S_1(0))}{Y_1(K_1+S_1(0))^5} \\
\frac{-(1-Y_1)\mu_{max_2}X^4S_1(0)K_{m_1}^2S_1(0)-8K_{m_1}S_1(0)+6S_1(0)^2)/K_{m_1}}{Y_1(K_1+S_1(0))^7} \\
\vdots
\end{bmatrix}. $$

A reparameterisation is $s(\theta) = [(1 - Y_1)\mu_{max_1}XS_1(0)/Y, K_{m_1} + S_1(0), -K_{m_1}\mu_{max_1}X/Y, (K_{m_1}-2S_1(0))/K_{m_1}, (K_{m_1}^2-8K_{m_1}S_1(0)+6S_1(0)^2)/K_{m_1}, \ldots]$. However, this is an infinite reparameterisation and by Remark 7 it is not possible to use Theorem 8b. Instead we suggest using the finite reparameterisation $s(\theta) = [s_1, s_2, s_3, s_4]^T = [(1 - Y_1), \mu_{max_1}X/Y, K_{m_1} + S_1(0), S_1(0)]^T$. Rewriting the exhaustive summary in terms of $s$ gives

$$\kappa(s) = \begin{bmatrix}
\frac{s_1s_2s_4}{s_3} \\
\frac{s_1s_2s_4(-s_3 + s_4)/s_3^3} \\
\frac{s_1s_2s_4(-s_3^2 - 4s_3s_4 + 3s_4^2)/s_3^5} \\
\frac{s_1s_2s_4(-s_3^3 + 11s_3^2s_4 - 25s_3s_4^2 + 15s_4^3)/s_3^7} \\
\vdots
\end{bmatrix}. $$

The rank of $D_s = [\partial\kappa_j/\partial s_i]$ is $r_s = 3$ (see Example6.mw), so that the reparameterisation is not a reduced-form exhaustive summary. By Theorem 7b, the number of estimable parameters in the model is 3. Therefore, the single monod model is parameter redundant with deficiency 2. Further, by solving the PDEs of equation (6), we can show that a reduced-form exhaustive summary is

$$r(\theta) = \left[\frac{\mu_{max_1}X(1 - Y_1)}{Y_1}, (K_{m_1} + S_1(0))(1 - Y_1), S_1(0)(1 - Y_1)\right], $$

which also forms a set of estimable parameters, as given in [46].

Next we consider the double monod model. The obvious exhaustive sum-
mary is again the coefficients of a Taylor-series expansion of equation (11) with

\[
\kappa(\theta) = \begin{bmatrix}
  \frac{(1-Y_1)\mu_{\max 1}X_1(0)}{Y_1(K_{m1}+S_1(0))} + \frac{(1-Y_2)\mu_{\max 2}X_2(0)}{Y_2(K_{m2}+S_2(0))} \\
  \frac{-\mu_{\max 1}X_1^2S_1(0)K_{m1}}{Y_1^2(K_{m1}+S_1(0))^2} + \frac{-\mu_{\max 2}X_2^2S_2(0)K_{m2}}{Y_2^2(K_{m2}+S_2(0))^2} \\
  \vdots
\end{bmatrix}\]

With a similar finite reparameterisation to that in the single monod case, \(s(\theta) = [(1 - Y_1), \mu_{\max 1}X/Y_1, K_{m1} + S_1(0), S_1(0), (1 - Y_2), \mu_{\max 2}X/Y_2, K_{m2} + S_2(0), S_2(0)]^T\), the rank of \(D_s = [\partial \kappa_1/\partial s]\) is \(r_s = 6\), so that the reparameterisation is not a reduced-form exhaustive summary. By Theorem 7b, the number of estimable parameters in the model is 6. Therefore, the double monod model is parameter redundant with deficiency 3. Again by solving equation (6), we can show that a reduced-form exhaustive summary is \(r(\theta) = [\mu_{\max 1}X(1 - Y_1)/Y_1, (K_{m1} + S_1(0))(1 - Y_1), S_1(0)(1 - Y_1), \mu_{\max 2}X(1 - Y_2)/Y_2, (K_{m2} + S_2(0))(1 - Y_2), S_2(0)(1 - Y_2)]^T\), which again also forms a set of estimable parameters, as also given in [46].

Alternatively the extension theorem (Theorem 3) could be applied using Remark 1, as we are considering a parameter-redundant model. First we split the exhaustive summary into a new exhaustive summary of the form

\[
\kappa(\theta) = \begin{bmatrix}
  \kappa_1 \\
  \kappa_2
\end{bmatrix}
= \begin{bmatrix}
  \frac{(1-Y_1)\mu_{\max 1}X_1(0)}{Y_1(K_{m1}+S_1(0))} \\
  \frac{-\mu_{\max 1}X_1^2S_1(0)K_{m1}}{Y_1^2(K_{m1}+S_1(0))^2} \\
  \vdots
\end{bmatrix}
= \begin{bmatrix}
  s_1, s_1, s_2, \ldots, s_{1,1}, s_{1,2}, \ldots, s_{1,3} \\
  s_1, s_2, \ldots, s_{2,1}, s_{2,2}, \ldots, s_{2,3}
\end{bmatrix}
\]

with \(s = [s_1, s_2]\), where \(s_1 = [s_{1,1}, s_{1,2}, s_{1,3}] = [\mu_{\max 1}X(1 - Y_1)/Y_1, (K_{m1} + S_1(0))(1 - Y_1), S_1(0)(1 - Y_1)]\) and \(s_2 = [s_{2,1}, s_{2,2}, s_{2,3}] = [\mu_{\max 2}X(1 - Y_2)/Y_2, (K_{m2} + S_2(0))(1 - Y_2), S_2(0)(1 - Y_2)]\). Then, as \(D_1 = [\partial \kappa_1/\partial s]\) is full rank and \(D_{2,2} = [\partial \kappa_2/\partial s]\) is also full rank, by Theorem 3 \(s\) is a reduced-form exhaustive summary. Therefore, by the reparameterisation theorem we can show again
that the double monod model is parameter redundant and 6 parameters are estimable.

Using induction we can extend this to \( k \) of any size. The model with parameters \( \theta = [Y_1, S_1(0), \mu_{\text{max}1}, K_{m1}, \ldots, Y_k, S_k(0), \mu_{\text{max}k}, K_{mk}, X] \) is parameter redundant with deficiency \( k + 1 \). The \( 3k \) estimable parameters are of the form \([\mu_{\text{max}}i X(1 - Y_i)/Y_i, (K_{mi} + S_i(0))(1 - Y_i), S_i(0)(1 - Y_i)]\) for \( i = 1, \ldots, k \). This illustrates one of the big advantages of the exhaustive summary framework, the ability of find general results. [46] was able to derive with a set of estimable parameters for the model with \( k = 1 \), but this is not extendable to any \( k \); the exhaustive summary framework allows for this.

4.3. Example 7 - Naive Bayesian Networks

A naive Bayesian network consists of \( n \) observable nodes, \( Y_1, \ldots, Y_n \), and a single unobservable node, \( Z \). In this example all the nodes are binary (0 or 1). The model has \( 2n + 1 \) parameters \( \theta = [p, \theta_{1|1}, \ldots, \theta_{n|1}, \theta_{1|0}, \ldots, \theta_{n|0}] \), where \( p \) is the probability that \( Z \) equals 1 and \( \theta_{i|j} \) is the probability that \( Y_i \) equals 1 given that \( Z \) equals \( j \). The probability of an observation \( y \) is given by

\[
P(y) = p \prod_{i=1}^{n} \theta_{i|1}^{y_i} \bar{\theta}_{i|1}^{1-y_i} + \bar{p} \prod_{i=1}^{n} \theta_{i|0}^{y_i} \bar{\theta}_{i|0}^{1-y_i}.
\]

This provides an exhaustive summary

\[
\kappa = \begin{bmatrix}
p \prod_{i=1}^{n} \bar{\theta}_{i|1} + \bar{p} \prod_{i=1}^{n} \bar{\theta}_{i|0} \\
p \prod_{i=1}^{n-1} \bar{\theta}_{i|1} \theta_{n|1} + (1 - p) \prod_{i=1}^{n-1} \bar{\theta}_{i|0} \theta_{n|0} \\
\vdots \\
p \prod_{i=1}^{n} \bar{\theta}_{i|1} + \bar{p} \prod_{i=1}^{n} \bar{\theta}_{i|0}
\end{bmatrix}.
\]

(As all these terms sum to 1, it is also possible to exclude one term from the exhaustive summary.)

A Jacobian matrix is used by [47, 48] to show that when \( n \) equals 3 this model is full rank (the model is parameter redundant with deficiency 2 when \( n = 2 \)). [47] points out that the model is only conditionally full rank, and presents 7 non-identifiable models. These non-identifiable models can easily be found using
Theorem 4. The Maple worksheet Example7.mw presents the calculation of the derivative matrix for the case $n = 3$, which involves differentiating the terms of equation (12) with respect to the parameters $\theta = [p, \theta_{1|1}, \theta_{2|1}, \theta_{3|1}, \theta_{1|0}, \theta_{2|0}, \theta_{3|0}]$. The rank of the derivative matrix is 7, so that this model is full rank. A PLUR decomposition of the derivative matrix results in the matrix $U$ having a determinant of $\text{Det}(U) = -p^3\bar{p}^3(\theta_{1|1} - \theta_{1|0})^2(\theta_{2|1} - \theta_{2|0})^2(\theta_{3|1} - \theta_{3|0})^2$. This model is parameter redundant if $\theta_{1|1} = \theta_{1|0}$ or $\theta_{2|1} = \theta_{2|0}$ or $\theta_{3|1} = \theta_{3|0}$. Therefore, this model would be parameter redundant if $\theta_{1|1} = \theta_{1|0} = 0$, so that there was no flow from node $Z$ to node $Y_1$, or any other combination with no flow to the observable nodes, as shown in [47]. [48] conjecture that only these points in the parameter space lead to non-identifiability. However this conjecture is not quite true as any points with $\theta_{1|1} = \theta_{1|0}$ or $\theta_{2|1} = \theta_{2|0}$ or $\theta_{3|1} = \theta_{3|0}$ can lead to non-identifiability.

However, when $n = 4$ it is not possible to calculate the rank of the derivative matrix, and the extension theorem (Theorem 3) cannot be applied to the exhaustive summary given by equation (12), because adding the additional parameters $\theta_{4|1}$ and $\theta_{4|0}$ changes this exhaustive summary. [47] uses a method specific to naive binary networks to show that this model is also full rank. Instead we can consider a reparameterisation where the extension theorem may be applied.

A reparameterisation that would allow the use of the extension theorem is $s = [p\theta_{1|1}\theta_{2|1}\theta_{3|1}, p\bar{p}\theta_{1|1}\theta_{2|1}, p\theta_{1|1}\theta_{2|0}\theta_{3|0}, p\bar{p}\theta_{1|0}\theta_{2|0}, p\bar{p}\theta_{1|0}, p]^T$. The derivative matrix $D_s = [\partial \kappa_{s,j}/\partial s_i]$ is full rank, and therefore $s$ is a reduced-form exhaustive summary.

Using Theorem 7b we come to the same conclusion that the model is full rank. If we extend the model to include an extra observable node, $y_4$, this introduces two extra parameters $\theta_+ = [\theta_{4|1}, \theta_{4|0}]$ and extra exhaustive summary terms, $s_+ = [p\theta_{1|1}\theta_{2|1}\theta_{3|1}\theta_{4|1}, (1 - p)\theta_{1|1}\theta_{2|0}\theta_{3|0}\theta_{4|0}]^T$. The matrix

$$D_+ = \begin{bmatrix} \partial \kappa_{s,+}/\partial s_+ \end{bmatrix} = \begin{bmatrix} p\theta_{1|1}\theta_{2|1}\theta_{3|1} & 0 \\ 0 & (1 - p)\theta_{1|1}\theta_{2|0}\theta_{3|0} \end{bmatrix}$$
is of full rank, and then, by Theorem 3, the model is also full rank when \( n = 4 \). By induction the binary nodes example will always be full rank when \( n \geq 3 \).

This conclusion was reached by [47], but using a method specific to this particular type of model. The present method has the advantage that any points that are parameter redundant within this full rank model can be found using a PLUR decomposition.

5. **Inference, missing data and weak identifiability**

It is shown in [5] that deterministic models such as compartment models and that of example 6 can be included in the exponential family by suitable addition of an error structure. In that paper it is also shown that in the exponential family case parameter redundancy equates to having a ridge in the likelihood surface. Such a ridge then naturally precludes unique parameter estimation by maximum likelihood. If \( M(\boldsymbol{\theta}) \) consists of the terms of a log-likelihood, then global identifiability will imply uni-modality of the log-likelihood surface.

By focussing on model structure, the work of this paper essentially assumes complete data. However, missing data can effectively turn a full-rank model into a parameter-redundant one, by removing terms from the exhaustive summaries used to describe the model. In this case an exhaustive summary that includes the data can be used, such as the terms that make up the log-likelihood.

*Example 1 CJS Model Continued.* In this example with specific data we
could use the exhaustive summary

\[
\kappa = \begin{bmatrix}
  n_{1,2}\ln(\phi_1p_2) \\
  n_{1,3}\ln(\phi_1\phi_2p_2p_3) \\
  n_{1,4}\ln(\phi_1\phi_2\phi_3p_2p_4) \\
  F_1\ln\{(1 - \phi_1p_1 - \phi_1\phi_2p_2p_3 - \phi_1\phi_2\phi_3\bar{p}_2p_4)\} \\
  n_{2,3}\ln(\phi_2p_3) \\
  n_{2,4}\ln(\phi_2\phi_3p_3p_4) \\
  F_2\ln\{(1 - \phi_2p_3 - \phi_2\phi_3\bar{p}_3p_4)\} \\
  n_{3,4}\ln(\phi_3p_4) \\
  F_3\ln\{(1 - \phi_3p_4)\}
\end{bmatrix},
\]

where \(n_{i,j}\) is the number of animals ringed at time \(i\) and first seen again at \(j\) and \(F_i\) is the number of animals ringed at time \(i\) but never seen again. If \(n_{1,3} = n_{1,4} = 0\) the rank of \(D = [\partial\kappa_k/\partial\theta_i]\) is still 5. However, if \(n_{1,2} = n_{1,3} = n_{1,4} = 0\) the rank of \(D = [\partial\kappa_k/\partial\theta_i]\) is 4, so that in this case of missing data there are now only 4 estimable parameters in this model.

For any particular modelling application, it is interesting to consider how much data could be missing before the identifiability status of the model changes, and this is work in progress.

In a Bayesian context, the presence of a ridge in a likelihood surface can still result in a posterior distribution with precise estimates, even if uniform prior distributions are employed for all of the model parameters. An illustration of this is provided in [49], where use is made of an explicit maximum likelihood estimate of the parameters of a model for ring-recovery data obtained from records of dead marked animals [50]. In this case it can be shown that the known orientation of the ridge in the likelihood surface is such that precise marginal posterior distributions may result even when a model is parameter redundant and the prior distributions are flat. Dangers of poor mixing and slow convergence of MCMC samplers in such cases are well known; see for example page 187 of [51].

The Bayesian equivalent to near-redundancy is weak identifiability, defined
by [52] as occurring when particular marginal prior and posterior distributions are similar. In this case the corresponding parameter is said to be weakly identifiable. [53] advocate examining univariate marginal prior/posterior overlap probabilities, and suggest an ad hoc 35% guideline, with overlap greater than that value indicating weakly identified parameters. This criterion is examined by [54] for capture-recapture models and is found to be effective when uniform prior distributions are used.

6. Discussion

In this paper we have devised and used new computational tools for determining the parametric structure of models, focussing in particular on the use of symbolic algebra procedures. However, in some cases it is possible to determine analytically the redundancy status of models, without recourse to computation. Illustrations of this are provided by [41].

The examples of this paper have shown that the parametric structure of a model is not always as obvious as in example 1, where the parameters $\phi_3$ and $p_4$ only ever appear as a product. We have therefore developed a global framework for determining the parametric structure of models. The way in which this framework works is demonstrated in the diagram of Figure 1.

The key to determining the parametric structure of a model is the exhaustive summary. As an exhaustive summary only has to define a model uniquely, there are often many options for exhaustive summaries of various degrees of complexity, ranging from an original exhaustive summary that closely matches the model to the set of estimable parameters. The set of estimable parameters is the simplest form an exhaustive summary can take, but to be able to determine the set of estimable parameters we first need another exhaustive summary as a starting point. To determine whether a model is full rank or parameter redundant, the first stage is to determine the original exhaustive summary that is simplest in structure. For example, in a linear compartment model the transfer function exhaustive summary is structurally simpler than the Taylor-series exhaustive
Figure 1: A diagram showing the framework of the exhaustive summary approach
summary. A derivative matrix can then be formed from the chosen exhaustive summary, and the rank-derivative test can be used to determine whether or not a model is parameter redundant (Theorem 2). However, in structurally more complex models, such as example 5, the transfer function exhaustive summary results in a derivative matrix for which a computer runs out of memory trying to calculate the rank. Instead in such cases we need to use the reparameterisation theorem (Theorem 8) in order to seek a simpler reduced-form exhaustive summary.

Other tools exist to extend results. The extension theorem can be used to generalise a result for a class of models by starting with the smallest case of that model, as demonstrated in examples 1, 3, 6 and 7. For a full rank model, a PLUR decomposition can identify sub-models that are parameter redundant or any points in the parameter space at which a model is parameter redundant, or near parameter redundant as demonstrated in examples 3, 4 and 7.

The derivative matrix rank test can distinguish between full rank and parameter redundant models, and the PLUR decomposition can identify essentially full rank models. However, in order to determine if a model is locally or globally identifiable, Theorem 9 can be applied, which involves solving a set of equations. For this set of equations to be able to be solved they again need to be structural simple; therefore we suggest using a structurally simple reduced-form exhaustive summary in this case, as demonstrated in examples 3 and 5. We note again here that the distinction between essentially full rank and conditionally full rank, while convenient mathematically, can reduce to differences in specification of the parameter space. A model which is conditionally full rank can become essentially full rank by excluding small numbers of singular points or regions of the parameter space. This then allows Theorem 9 to be applied to conditionally full rank models, and we have seen an instance of this in our discussion of example 3.

When the rank-derivative test has failed to work, a numerical method may be used instead, see for example [44]. However, a numerical method can only tell us the approximate number of estimable parameters at a single point in
the parameter space. The exhaustive summary approach gives the definitive answer to whether a model is parameter redundant or not, and exactly what is estimable. It can be easily extended to the general case using the extension theorem. It is also relatively easy to establish the parameter redundancy status of sub-models of a full rank model. Alternatively the structurally reduced-form exhaustive summary can be differentiated with constraints applied with a simple loop code in Maple as in example 4. This exhaustive summary framework also removes the need to find a model-specific proof that a model is full rank, as in [47].

The exhaustive summary framework of this paper is perfectly general. In theory it could work for any parametric model. So far the application of the reparameterisation method has worked in every complex model we have seen, often where either previous methods have failed or a model-specific method for determining parametric structure has been used instead.

References


Appendix A. Proofs of theorems

We make use of three matrix results. Let \( A = BC \).

(i) From Theorem 1.6 of [55], if \( B \) is a square non-singular matrix, then \( \text{rank}(A) = \text{rank}(C) \).

(ii) From Section 0.4.5 of [56] if \( C \) is an \( n \times m \) matrix with rank \( n \), then \( \text{rank}(A) = \text{rank}(B) \).

(iii) From Section 0.4.5 of [56] if \( B \) is an \( \ell \times n \) matrix of rank \( n \), then \( \text{rank}(A) = \text{rank}(C) \).
Appendix A.1. Theorem 4

Both $P$ and $L$ are non-singular, by construction. Suppose that $\text{Det}(U) \neq 0$ at $\theta$. Then, at $\theta$, $PLU$ is non-singular, and so $\text{rank}(D) = \text{rank}(R) = q$. As $R$ is in reduced echelon form, then $\text{rank}(R) = q$ for all values of $\theta$. As $R$ is of full rank, then $\text{rank}(D) = \text{rank}(PLU)$. However, if $\text{Det}(U) = 0$ at $\theta$, $U$ is not full rank, so that $\text{rank}(PLU) = \text{rank}(U) < q$, and $\text{rank}(D) < q$ at $\theta$.

Appendix A.2. Theorem 5

In what follows, we assume that $R$ is defined. Let $z$ denote the points in the parameter space at which $\theta_j = \theta_i$. Suppose $\text{Det}(U) = 0$ at $z$. Let $D^*$ be $D$ evaluated at $z$ and let $D^* = P^*L^*U^*R^*$. Then $\text{rank}(P^*L^*U^*) = \text{rank}(P^*L^*U^*)$ and $\text{rank}(P^*L^*U^*) = \text{rank}(U^*)$, so that $\text{rank}(D^*) = \text{rank}(U^*)$.

Let $D_{\text{sub}}$ be the derivative matrix of a sub-model of the full rank model at $z$. Then $D_{\text{sub}} = ED^*$, for some $E$.

Let $\text{rank}(U^*) = r_U < p$. There are then $d_U = p - r_U$ solutions to $(\alpha^*)^T D^* = 0$. Label the solutions, $\alpha_{m}^*$ for $m = 1, \ldots, d_U$, with individual entries $\alpha_{m,k}^*$, for $k = 1, \ldots, p$. As $\theta_j = \theta_i$, each $\alpha_{m,j}^* = \alpha_{m,i}^*$. Let $\alpha_{m}'$ be $\alpha_{m}^*$ without entry $\alpha_{m,j}^*$. Then $(\alpha_{m}^*)^T = (\alpha_{m}'^*)^T E$. As

$$(\alpha^*)^T D^* = 0 \Leftrightarrow (\alpha')^T ED^* = 0 \Leftrightarrow (\alpha')^T D_{\text{sub}} = 0$$

and there are $d_U$ solutions, the deficiency of $D_{\text{sub}}$ is $d_U$ so that the sub-model is parameter redundant, with deficiency equal to the deficiency of $U^*$. If the sub-model is not parameter redundant, then

$$(\alpha')^T D_{\text{sub}} = 0 \Rightarrow (\alpha')^T = 0$$
$$\Leftrightarrow (\alpha')^T ED^* = 0 \Rightarrow (\alpha')^T = 0$$
$$\Leftrightarrow (\alpha^*)^T D^* = 0 \Rightarrow (\alpha^*)^T = 0.$$

Therefore $D^*$ has full rank, which will only occur if $\text{Det}(U) \neq 0$ at $z$. The extension to more complex linear constraints on the elements of $\theta$ is immediate.
Appendix A.3. Theorem 8

a. Either $s = r$, of reduced form, or there exists a reparameterisation $r$ which has $\text{dim}(r) = r_s$ and $\text{rank}(D_r) = r_s$:

$$D_\theta = \frac{\partial \kappa}{\partial \theta} = \frac{\partial x}{\partial \theta} \frac{\partial \kappa}{\partial r} = D_2 D_r.$$ 

Because $D_r$ is an $r_s \times n$ matrix with rank $r_s$, the rank of $D_2 D_r$ is the same as the rank of $D_2$.

b.

$$D_\theta = \frac{\partial \kappa}{\partial \theta} = \frac{\partial s}{\partial \theta} \frac{\partial \kappa}{\partial s},$$

where $(\partial s/\partial \theta)$ is a $p \times p_s$ matrix with rank $p_s$. Therefore, $\text{rank}(D_\theta) = \text{rank}(\partial \kappa/\partial s) = r_s.$