The Evaluation of Evidence for Autocorrelated Data in Relation to Traces of Cocaine on Banknotes

Amy Wilson
School of Mathematics and Maxwell Institute, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, UK.

Colin Aitken
School of Mathematics and Maxwell Institute, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, UK.
c.g.g.aitken@ed.ac.uk

Richard Sleeman
MSA Ltd., Building 20F, Golf Course Lane, P.O. Box 77, Filton Road, Bristol, BS34 7RP, UK

James Carter
Queensland Health Forensic and Scientific Service, 39 Kessels Road, Coopers Plains, QLD 4109, Australia

Summary.
Much research in recent years for evidence evaluation in forensic science has focused on methods for determining the likelihood ratio in various scenarios. When the issue in question is whether evidence is associated with a person who is or is not associated with criminal activity then the problem is one of discrimination. A procedure for the determination of the likelihood ratio is developed when the evidential data are believed to be driven by an underlying latent Markov chain. Three other models that assume autocorrelated data without the underlying Markov chain are also described. The performances of these four models and a model assuming independence are compared using data concerning traces of cocaine on banknotes.

Keywords: autoregressive model, cocaine on banknotes, evidence evaluation, hidden Markov model, likelihood ratio, nonparametric density estimation.

1. Introduction

1.1. The use of drug traces on banknotes as evidence
In a criminal case, evidence $E$ is often evaluated through use of a likelihood ratio (LR), the factor which converts the prior odds in favour of a proposition proposed by the prosecution, as compared with a proposition proposed by the defence, to the posterior odds. The word ‘odds’ is used advisedly; the propositions do not need to be exhaustive. The LR is known as the value of the evidence. A statistical approach is of particular interest when the evidence is in the form of data. There are many applications where such data are multivariate continuous and independent, such as elemental compositions of glass fragments or chemical compositions of drugs (see Aitken and Taroni, 2004, for examples). In the example described
here the data are univariate and autocorrelated. Various models are proposed for the
evaluation of such data and their performance assessed.

The data are the quantity of cocaine on each of a sample of banknotes as measured
by a mass spectrometer. The quantity of cocaine on a banknote in a sample of banknotes
is associated with the quantity of cocaine on adjacent banknotes in that sample. There
has been some previous use of LRs in the area of drugs on banknotes. In Besson (2004)
the likelihood ratio of the quantity of contamination of cocaine on a seized banknote is
evaluated using a histogram. It is noted that calculating a LR for a sample consisting
of multiple banknotes using this method is not possible without assuming independence.
In Taroni et al. (2010) the likelihood ratio for the quantity of cocaine contamination on
each of a sample of banknotes is calculated using a univariate kernel density estimate.
An assumption of independence is made and it is noted that this assumption may not be
warranted. This assumption is not made in the models introduced in this paper.

Methods of evidence evaluation for drugs on banknotes that do not use the LR approach
are studied in Ebejer et al. (2005), Dixon et al. (2006), Lloyd (2009) and Jourdan et al.
(2013). An arcsin (square root) and a log transformation are both used in Ebejer et al.
(2005) to model the distribution of the proportion of banknotes contaminated with diamor-
phine for samples of banknotes from general circulation. A one sample t-test is carried out
to compare the percentage of contamination found in a sample of banknotes seized by law
enforcement agencies to each of the distributions fitted to general circulation banknotes. If
result of this t-test is significant, then it is concluded that the sample of seized banknotes are
unlike general circulation. This approach uses only information as to whether a banknote is
contaminated or not and does not account for the quantity of contamination. Diamorphine
contamination on banknotes is infrequent so consideration of the proportion contaminated
in a seizure may be sufficient as evidence of criminal association. Jourdan et al. (2013) use
a similar approach. A Pareto distribution is fitted to the quantities of cocaine contamina-
tion on banknotes from general circulation. The quantity of cocaine contamination found
on a banknote seized by a law enforcement agency is said to be unusual if the probability
of obtaining an interval of contamination in which the seized banknote falls is unlikely, as
determined by the fitted Pareto distribution. Dixon et al. (2006) and Lloyd (2009) use
principal components analysis to reduce the measurements of the quantity of cocaine on
each of a sample of n banknotes to a small number of principal components. Classification
methods, including the Mahalanobis distance and support vector data description are then
used to classify a sample of banknotes seized by law enforcement agencies as being either
from general circulation or not (Lloyd, 2009) or from general circulation or associated with
crime (Dixon et al., 2006).

The role of likelihood ratios for the analysis of data of a hierarchical random effects
nature in forensic science was introduced by Lindley (1977). One advantage of using a
likelihood ratio instead of a traditional hypothesis test is that it has no dependence on an
arbitrary cut-off point (e.g., 5% significance). Also an approach based on the likelihood
ratio ensures equality of treatment of both propositions in contrast to hypothesis testing. A
discussion of the merits of the LR approach can be seen in Aitken and Taroni (2004). In this
paper, a LR approach is used to evaluate evidence relating to traces of drugs on banknotes,
and no assumption of independence between the quantity of cocaine on adjacent banknotes
is made. Instead, there is an assumption of autocorrelation at lag one. The results obtained
from the models discussed here are compared with the results from a model which assumes
The Evaluation of Evidence for Autocorrelated Data in Relation to Traces of Cocaine on Banknotes

The banknotes on which the quantities of cocaine are measured in order to determine the value of the evidence (LR) are provided by law enforcement agencies to an analytical chemistry laboratory. The laboratory measures the quantity of cocaine either on all of the banknotes or on a subset of the banknotes and is interested in the support provided by the results for one or other of the following propositions:

- \( H_C \): the banknotes are associated with a person who is associated with a criminal activity involving cocaine, and
- \( H_B \): the banknotes are associated with a person who is not associated with a criminal activity involving cocaine.

The data used for the analysis, \( z = (z_1, z_2, \ldots, z_n) \), known as the seized sample, are the logarithms of the peak areas (where the peaks are obtained using a mass spectrometer) corresponding to cocaine on a set of \( n \) banknotes. The strength of the evidence of \( z \) in support of \( H_C \) or \( H_B \) is to be assessed. Training data are available, first, from banknotes from cases that went to trial and in which the defendant was convicted (either by trial or through a plea of guilty) of a crime involving cocaine and, second, from banknotes associated with general or background circulation. These training data are used to develop models associated with \( H_C \) and \( H_B \), respectively.

The likelihood ratio \( LR \) associated with the propositions \( H_C \) and \( H_B \) is given by:

\[
LR = \frac{f(z | H_C)}{f(z | H_B)},
\]

where the function \( f \) is a probability density function for the measurements, conditional on \( H_C \) and on \( H_B \) in the numerator and the denominator, respectively. If this statistic is greater than one, then the evidence assigns more support to the proposition that the banknotes are associated with a person who is associated with drug crime involving cocaine relative to the proposition that the banknotes are associated with a person who is not associated with drug crime involving cocaine. With an assumption of independence amongst the values in \( z \),

\[
LR = \prod_{i=1}^{n} \frac{f(z_i | H_C)}{f(z_i | H_B)}.
\]

The interpretation of this LR is slightly different from the one used for comparison of possible sources of recovered and control evidence in Lindley (1977). Here there is only one set of evidence, the seized banknotes provided by the law enforcement agency. The LR provides a measure of support for one or other of the propositions as to whether the person with whom they are associated is himself associated or not with criminal activity involving cocaine.

A hidden Markov (HM) model is described which assumes that the data \( z \) have an underlying latent Markov chain. The states of the Markov chain refer to different levels of contamination within a sample of banknotes. The results from the HM model are compared with a model that assumes independence, known as the standard model and with three other models that account for autocorrelation of lag one but without the underlying
Markov structure, an autoregressive model and two non-parametric models, one with a fixed bandwidth and one with a variable bandwidth. The HM model enables an extension of these models to consider different levels of contamination (of which it is assumed there are two); these levels can arise because samples of banknotes which are associated with a crime involving cocaine may contain subsets of banknotes (known as bundles) which have different origins and hence may have different distributions of contamination. These different levels of contamination can also arise in samples of banknotes from general circulation which may, by chance, contain banknotes which are highly contaminated.

Training data arising from samples of banknotes that are associated with a person who was convicted of a crime involving cocaine may contain samples of banknotes which were not actually directly involved in such a crime (even though the suspect with whom they are associated was convicted). These banknotes may, instead, have been obtained by innocent means. As such, some samples in the training database associated with $H_C$ may be contaminated with low quantities of cocaine, such as may arise from being in general circulation. This possibility leads to a difficulty in the assessment of performance of the models which is discussed later.

The aim of the paper is to compare the performance of the five models for the evaluation of the likelihood ratio $f(z \mid H_C)/f(z \mid H_B)$. This will be done with reference to two sets of training data, one associated with people who have been involved in drug crime relating to cocaine and one associated with general circulation. Descriptions for forensic scientists of the standard model, the autoregressive model and the non-parametric models are given in Wilson et al. (2014). The full description of the HM model and a more technical description of the three models which use dependence are given here. Results are presented for all five models.

2. Data

Banknotes are analysed one at a time, using a mass spectrometer (as described in Dixon et al., 2006). The output is given by the number of gas phase ion transitions detected at each scan number, where ten scans occur every second. The ion used for discussion of the methods described here is the cocaine product ion $m/z$ 105. The output for a sample of banknotes consists of a series of peaks, with each peak corresponding to one banknote. The height of the peak at any given scan number is given by the gas phase ion transition count (for cocaine product ion $m/z$ 105). The areas of these peaks can be calculated with the aid of a peak detection algorithm, details of which are available from the corresponding author. The logarithms of the peak areas, a measure of the amount of cocaine on each banknote, are used as the data for analysis. Logarithms of the peak areas are used to reduce skewness and to ensure consistency with earlier work such as Dixon et al. (2006) and Lloyd (2009). The methods may also be applied to other ions and future work will consider the multivariate problem of analysing the data from the two ions on each of five drugs.

Banknotes that are associated with a person who has been involved with crime

A training dataset, $C$, was compiled of samples of banknotes that had been involved in a criminal case in which the defendant was convicted of a crime involving cocaine. A single criminal case may consist of multiple samples of banknotes, known as exhibits. For future
The Evaluation of Evidence for Autocorrelated Data in Relation to Traces of Cocaine on Banknotes

reference, any sample of banknotes in \( C \) will be known as an exhibit. The law enforcement agencies decide on seizure how banknotes associated with one criminal case are divided into individual exhibits. This decision, for example, could be based on the location in which the banknotes were found. The training data in \( C \) are denoted \( y = \{ y_{it}; i = 1, \ldots, n^C, t = 1, \ldots, n^C_i \} \); the logarithms of the peak areas for cocaine; there are \( n_C \) exhibits with \( n^C_i \) notes in the \( i \)-th exhibit. There are 29 cases containing at least one exhibit with greater than 20 banknotes. The 29 cases consist of between one and six exhibits, and there are a total of 70 exhibits which are known to have been associated with a person who has been involved in drug crime relating to cocaine. The definition of set \( C \) used here is different from previous definitions used in Besson (2004) and Dixon et al (2006), for example, when all banknotes provided by law enforcement agencies were used as training data.

There are two difficulties with the definition of set \( C \). The first is that the evidence provided from the analysis may have influenced the outcome of the trial (and so biased the data). In fact, just twelve of the 70 crime exhibits were declared as contaminated by the experts, suggesting that this is not too much of an issue. However, this leads to another problem, in that just because a suspect was convicted, it does not mean that the banknotes in the exhibits associated with that suspect were contaminated. These difficulties were further exacerbated by the fact that whilst the result of a criminal case might be known to be a conviction, it is not possible to map this result onto individual exhibits of banknotes. As a result, a large number of the exhibits in set \( C \) appear to consist of banknotes that have only low-level or background level contamination. Therefore, when the performances of the models are tested, it is expected that many of the exhibits in set \( C \) will falsely provide support (in the sense of having a LR < 1 when treated as the seized sample \( z \)) for proposition \( H_B \).

General Circulation Banknotes

Let set \( B \) be the training dataset consisting of samples of banknotes from general circulation (also known as background). The training data in \( B \) are denoted \( x = \{ x_{it}; i = 1, \ldots, m_B, t = 1, \ldots, n^B \} \); the logarithms of the peak areas for cocaine; there are \( m_B \) samples of banknotes with \( n^B_i \) banknotes in the \( i \)-th sample. 193 samples of banknotes of English or Scottish currency were obtained from a variety of locations around the UK. For future reference, any set of banknotes used in the analyses discussed that is in set \( B \) will be known as a sample. This is in contrast to exhibit, which is used to describe banknotes in set \( C \).

A large number of samples of banknotes from general circulation were taken from the Bristol area with other samples being selected from various locations in England, Wales and Scotland. There have been some tests (Ebejer, 2007b) carried out which did not find evidence that the region that banknotes in general circulation came from has an effect on the quantities of drug found, but for the calculation of likelihood ratios for seized banknotes it may be necessary to tailor the background database to the region in which the crime occurred (though this is not an issue that is discussed further here). In addition, the majority of samples from general circulation used for analysis were taken from banks. This, again, could be a problem that requires further investigation if the defendant maintained that the banknotes had been acquired from some other provider.
2.1. Exploratory Data Analysis

Various features of the data in training sets B and C are discussed below.

Contamination on notes from general circulation

In previous studies of the quantities of cocaine on banknotes, (Besson, 2004, Dixon et al., 2006), it has been noted that using the percentage of contaminated banknotes within a sample as a statistic to distinguish between general circulation samples and crime exhibits is not possible due to the high frequency of contamination within samples from general circulation. A study of currency in general circulation in the USA demonstrated quantifiable levels of cocaine residue on approximately 97% of 418 10-note samples examined (Jourdan et al., 2013). It is therefore not sufficient to focus on the proportion of contaminated banknotes: the quantity of contamination needs to be taken into account to differentiate between crime exhibits and general circulation samples. A density plot (Wilson et al., 2014) of the mean quantity of cocaine contamination on 193 general circulation samples compared with a density plot of the mean contamination of the 70 crime exhibits where the means are determined from the individual quantities on the notes in the separate samples and exhibits shows that general circulation samples have mean quantities of contamination which, although generally lower than those of the crime exhibits, are still high, and there is a substantial overlap between the two density plots. In addition, the means of the crime exhibits have a mode in a similar position (log contamination of 6.5) to the main mode of the means of the general circulation samples, as well as a second, greater, mode at a higher quantity of the log contamination at 7.2. This is further evidence to support the suggestion arising from consideration of court cases and verdicts in comparison to expert witness reports that the so-called crime exhibits contain a large number of exhibits which are only contaminated at a background level.

Correlation

Previous experiments (Ebejer et al., 2007a) indicated that it was possible for drug traces to pass from one contaminated banknote to an adjacent one. Banknotes in samples and in smaller exhibits are usually analysed in full and in the order in which they were stored. Banknotes in larger exhibits are usually analysed in pre-selected groups of banknotes, where the banknotes in each group were adjacent in the exhibit. As a result, any transfer of drug that had occurred between adjacent banknotes in the sample or exhibit, as discussed in Ebejer et al. (2007a), would result in autocorrelation being present within the analysed samples or exhibits. In addition, when banknotes are analysed, there is often no definitive end to the peak of ion counts obtained using the mass spectrometer. Some of the ion counts occurring from cocaine on the previous banknote may be included in the reading for the next peak. This effect could also result in autocorrelated data.

About 90% of the 193 samples from general circulation and 80% of the 70 crime exhibits had a significant sample partial autocorrelation coefficient of lag one; i.e., the observed partial autocorrelation coefficient of lag one is significantly different from zero at the 5% level (Chatfield, 2004, p. 56). The proportions of samples and exhibits with significant partial autocorrelation coefficients of higher order dropped to around 24% and 14% for
samples and exhibits, respectively, at lag two, and 6% and 7% by lag five. The parametric models described below fit an autoregressive process of order one to the data; models of higher order are not considered. The nonparametric models allow for dependence of the quantity of cocaine on a banknote on the quantity of cocaine on the previously analysed banknote.

_Different levels of contamination within samples and exhibits_

Figure 1 shows in the top-left plot a conditional density estimate (computed using the np package in R (Hayfield, 2008)) of the value of contamination of one banknote against the value of contamination of the previous banknote for one of the general circulation samples. It looks like a typical autoregressive process, with the cross-section at each value of the previous banknote being unimodal. When analysing similar plots of the crime exhibits, it was found that some of the exhibits did not have this unimodal distribution. Whilst some resembled an autoregressive process, others had many ridges and demonstrated some multimodal cross sections. This could reflect that a sample of banknotes taken from someone involved in drug crime may be a mixture of banknotes which are highly contaminated, and of banknotes which have just come from general circulation. For example, a drug dealer may take small bundles of banknotes from different customers (some of which may be contaminated from the customer’s drug use, and some of which may not), and will collect these together in large bundles. The large bundle would then consist of smaller bundles, some of which will have high levels of contamination, and some of which will have low levels of contamination. On the other hand, if a drug dealer were to contaminate the bundle evenly himself (say, by counting out banknotes onto a contaminated surface), or if he had obtained all of the banknotes from the same similarly contaminated source, the analysed sample might consist of similarly contaminated banknotes throughout. Some general circulation samples also exhibited multimodality in the cross sections of the conditional density functions, although these were fewer in number than for the crime exhibits.

Figure 1 also shows conditional density estimates of the contamination of banknotes given the contamination of the previous banknote for three of the crime exhibits. The bottom left plot, exhibit 8, is from an exhibit with 71 detected banknotes (detected in the sense of a peak being detected by the peak detection algorithm). The plot has two ridges, one large and one small. A banknote could therefore be followed either by a banknote with contamination commensurate with the larger ridge, or by a banknote with contamination in line with the smaller ridge. An autoregressive process with two mean levels, with the mean determined by a latent state could be used to model this behaviour. The top right plot (exhibit 4) shows the conditional density plot of another exhibit, this time with 132 banknotes. The relative weights of the ridges in this plot change as the contamination of the previous banknote changes. This plot has multimodal cross sections, as with exhibit 8, but the cross sections look very different when the previous banknote has low contamination, compared to when the previous banknote has high contamination. When the previous banknote has low contamination, the current banknote is most likely to be in the lower ridge, whereas when the previous banknote has high contamination, the current banknote is most likely to be in the highest ridge. This indicates that banknotes with low contamination are likely to be followed by other banknotes with low contamination, and that banknotes with high contamination are most likely to be followed by other banknotes with high contamination.
This dependence, of whether the banknote is ‘low’ or ‘high’, on the quantity of contamination of the previous banknote, could be modelled using an autoregressive process with mean determined by a latent state, where these states form a Markov Chain (i.e. a hidden Markov model), which would account for this dependence between contamination levels of adjacent banknotes for some of the exhibits. Contamination of this pattern would occur if the exhibit contained multiple bundles of banknotes from different origins, with banknotes within the same bundle remaining together within the larger collection of bundles, so that banknotes with similar levels of contamination are grouped together. Finally, the bottom right plot, exhibit 7, shows an exhibit with 276 detected banknotes. It shows just one main ridge, extending diagonally, which indicates that the contamination of this exhibit could be modelled with a more straightforward autoregressive process, with just one mean level.

Fig. 1. Conditional density plot of general circulation sample (top left) and three crime exhibits
3. Models

Four models are described here, the autoregressive model of lag one, the hidden Markov model, and the two nonparametric models, one with fixed and one with variable bandwidth. The results from these models are compared with the fifth model, the standard model that assumes independence between measurements on separate banknotes (Wilson et al., 2014).

3.1. Autoregressive models of order one

There are two models, one for exhibits of banknotes associated with a person who is associated with crime ($C$) and one for samples of banknotes associated with general circulation ($B$). For determination of the likelihood ratio, the form of the model is the same, only the parameters are different. The model is described with general notation for a particular sample or exhibit of banknotes from a general set $D$ (where $D$ should be replaced by $B$ or $C$ depending on which of the two sets is being considered). The data in this general set are given by $w_i$, with the $i$-th sample or exhibit in this general set given by $w_i = (w_{i1}, \ldots, w_{im_D})^T$, with $w_i$ substituting for $x_i$, $i = 1, \ldots, m_B$ or $y_i$, $i = 1, \ldots, m_C$, as appropriate.

An autoregressive model $AR(1)$ specifies the following relationship amongst the variables:

$$w_{i,t} - \mu = \alpha (w_{i,t-1} - \mu) + \epsilon_{it}$$

where $t = 2, \ldots, n_D^n$; $\epsilon_t \sim N(0, \sigma^2)$; $w_{i1} \sim N(\mu, \sigma^2)$ and $|\alpha| < 1$.

Prior and posterior distributions

The training data are used in conjunction with prior distributions for the model parameters to obtain draws from the posterior distribution of the parameters ($\mu, \sigma, \alpha$) of the autoregressive model. The marginal prior distributions for the parameters associated with the $i$-th general sample $w_i$ are given by:

- $\mu \sim N(\frac{1}{2}(\max(w_i) + \min(w_i)), \text{range}(w_i)^2)$;
- $\sigma^2 \sim \text{IG}(2.5, \beta)$, where IG denotes the inverse gamma distribution and $\beta$ is a hyperparameter; the form of the inverse gamma density function is given in the Appendix;
- $\beta \sim \Gamma(0.5, 4/\text{range}(w_i)^2)$.
- $\alpha \sim N(0, 0.25)$, with the result restricted to lie between -1 and 1.

These prior distributions are used to provide compatibility with the HM model of Section 3.2. This requirement for compatibility led in particular to the choice of the prior distributions and hyperparameters for the parameters $\sigma^2, \mu$ and $\beta$, which have the same form as those used in Richardson and Green (1997) and Rydén (2008). The hyperparameters for $\mu$ and $\beta$ are dependent on the data $w_i$. Ideally, the parameters of these prior distributions would instead be chosen by an expert and would not be dependent on the data. In Richardson and Green (1997) the choices of the priors and hyperparameters are described as “making only ‘minimal’ assumptions on the data”, a conclusion which is also
true here. The prior distribution of the parameter \( \mu \) is relatively uninformative over the range of sensible values for log peak areas; none of the samples or exhibits has a mean which is outside the range \((4.9, 8.0)\) and the range of each individual sample and exhibit is generally around two. Similarly, with a range of the data \( w_i \) of two, the prior distribution of \( \beta \) has most of its weight in the range \((0, 2)\). Values of \( \beta \) smaller than two result in prior distributions for \( \sigma^2 \) which do not place a large weight on values of \( \sigma^2 \) that are larger than around three. This is a reasonable assumption which is not very informative, given that all of the banknotes in both datasets have log peak areas that lie in the range \((2, 9)\). A truncated Normal prior is used for the autocorrelation parameter, as used in Albert and Chib (1993).

Draws from the posterior distribution of the parameters \((\mu, \sigma^2, \alpha, \beta)\), denoted \( \theta \), conditional on the data \( w_i \), can be obtained using a Metropolis-Hastings sampler (Wilson et al., 2014). Draws should be obtained separately for each of the general circulation samples and each of the crime exhibits. For the \( i \)-th sample in training set \( B \), denote \( \theta \) by \( \theta_{A_i}^B \), and for the \( i \)-th exhibit in training set \( C \) denote \( \theta \) by \( \theta_{A_i}^C \), where \( A \) denotes autoregressive.

### 3.2. Hidden Markov Models

In a hidden Markov model, each observed data point is associated with an unobserved state. The states form a Markov chain and determine the probability density functions of each data point conditional on the previous data point. A Markov switching model is used here, which also allows for dependence between adjacent data points, so that autocorrelation between the peak areas of adjacent banknotes is modelled. For some other examples of Markov switching models see Albert and Chib (1993) and McCulloch and Tsay (1993).

The states are used to model the contamination level of each banknote (either high or low) within a sample or exhibit. A state is associated with each banknote. As discussed in Section 2.1, exhibits of banknotes tend to contain multiple bundles of banknotes, each of which may have a different origin and hence these bundles may have different levels of contamination. Samples of banknotes from general circulation may contain some highly contaminated banknotes which have been in contact with traces of cocaine. The logarithm of the peak area associated with each banknote is modelled with a probability density function with one of four possible sets of parameter values. The particular set of parameters used is determined by the state of that banknote; the parameters correspond to the level of contamination of the banknote. The latent Markov property is required since banknotes within the same bundle (which is often secured with a rubber band or a paper strip) are adjacent to one another in the exhibit. Hence, the contamination level of a banknote depends on the contamination level of the previous banknote in the exhibit.

As for AR(1) models in Section 3.1, a general sample or exhibit \( w_i \) is considered. The states associated with this general sample or exhibit are given by \( S = (S_1, \ldots, S_n) \). The model is shown by the Bayesian network in Figure 2. As can be seen, the observations \( w_i \) are not conditionally independent given the latent variables \( S \).

The model has four hidden states, corresponding to two different levels of contamination (low and high). These states are given by:
The Evaluation of Evidence for Autocorrelated Data in Relation to Traces of Cocaine on Banknotes

Fig. 2. Bayesian network for the hidden Markov model. Nodes $W$ denote measurements on banknotes, nodes $S$ denote states of banknotes. Subscripts denote locations of the banknote in the exhibit or sample presented to the mass spectrometer.

<table>
<thead>
<tr>
<th>State (s)</th>
<th>Contamination level of previous note</th>
<th>Contamination level of current note</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>low</td>
<td>low</td>
</tr>
<tr>
<td>2</td>
<td>low</td>
<td>high</td>
</tr>
<tr>
<td>3</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>4</td>
<td>high</td>
<td>high</td>
</tr>
</tbody>
</table>

The transition matrix, giving the probabilities of moving between these states, is:

$$
P = \begin{pmatrix}
1 - p_{01} & p_{01} & 0 & 0 \\
0 & 0 & p_{10} & 1 - p_{10} \\
1 - p_{01} & p_{01} & 0 & 0 \\
0 & 0 & p_{10} & 1 - p_{10}
\end{pmatrix}
$$

Four states are used, so that the contamination level of both the current and previous banknote is described in each state. This means that an extra line of dependence, between each data point $w_{it}$ and the previous state $S_{t-1}$ is not required. This results in a transition matrix containing zeroes, as it is impossible to pass, for example, from a state with a current banknote with low contamination to a state with a previous banknote with high contamination.

The initial distribution of the hidden state of the first banknote is given by $(p_1, p_2, p_3, p_4)$.

The assumed probability model for $w_{it}$, conditional on the states $S$ is given by

$$w_{it} - \mu^{(1)}_{S_i} = \alpha(w_{i,t-1} - \mu^{(1)}_{S_{i-1}}) + \epsilon_i$$

where

$$\epsilon_i \sim N(0, \sigma_{S_i}^2), \text{ for } t \in \{2, \ldots, n_i^D\}, \ w_{i1} \sim N(\mu^{(1)}_{S_i}, \sigma_{S_i}^2)$$

and $S_t \in \{1, 2, 3, 4\}$. The subscript $S_t$ indicates that the parameter is dependent on the value of the $t$-th hidden state.
Only two parameters are required for each of the mean $\mu$ and variance $\sigma^2$, as there are only two different levels of contamination. Both the mean and the variance are allowed to vary between levels. Therefore, corresponding to the definition of the states given above, set $\mu_1^{(1)} = \mu_1^{(2)} = \mu_2^{(1)} = \mu_2^{(2)} = \mu_3^{(1)} = \mu_3^{(2)} = \mu_4^{(1)} = \mu_4^{(2)} = \mu_2$, $\sigma_1 = \sigma_3$ and $\sigma_2 = \sigma_4$. The parameters associated with each state, written in the order $(\mu_{S_t}^{(2)}, \mu_{S_t}^{(1)}, \sigma_{S_t}^2, \alpha)$ are therefore given by:

- State 1: $(\mu_1, \mu_1, \sigma_1^2, \alpha)$,
- State 2: $(\mu_1, \mu_2, \sigma_2^2, \alpha)$,
- State 3: $(\mu_2, \mu_1, \sigma_1^2, \alpha)$,
- State 4: $(\mu_2, \mu_2, \sigma_2^2, \alpha)$.

The model for each sample or exhibit is, therefore, specified by the parameters:

$$(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \alpha, p_{01}, p_{10}, p_1, p_2, p_3, p_4).$$

Prior and posterior distributions

Prior distributions of the same form as Section 3.1 are placed on the parameters $\mu_1$, $\mu_2$, $\sigma_1^2$, $\sigma_2^2$ and $\alpha$. The prior distributions of $\sigma_1^2$ and $\sigma_2^2$ are given hyperparameters, $\beta_1$ and $\beta_2$, with hyperpriors as for $\beta$ in section 3.1. The transition probabilities $p_{01}$ and $p_{10}$ are given the prior beta distribution $B(0.6, 4)$, truncated to lie between $2/n^D_i$ and $(n^D_i - 2)/n^D_i$, where $n^D_i$ is the number of banknotes in the $i$-th general sample. The initial distribution of the hidden states was taken to be fixed and known, setting $p_1 = p_2 = p_3 = p_4 = 0.25$ (Cappé et al., 2005, p. 478).

It was found that the posterior distribution of the parameters could be sensitive to the hyperparameters used in the prior distributions placed on the transition probabilities, especially when small transition probabilities were allowed. This issue is discussed in Gassiat and Rousseau (2013). The parameters in the prior distributions for $p_{01}$ and $p_{10}$ were chosen as in Spezia (2010), and the distributions were truncated so as not to allow small values. Prior distributions were tested on simulated data with similar properties to the real data before use, and adjusted based on results from these simulations.

The training data are used in combination with the prior distributions to estimate the following parameters, which we denote for brevity by $\theta$:

$$\theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \alpha, p_{01}, p_{10}, \beta_1, \beta_2).$$

The estimates of the parameters take the form of draws from the posterior distribution $f(\theta | w_i)$. A Metropolis-Hastings sampler can be used to obtain these draws, details of which are given in the Appendix. For the $i$-th sample in training set $B$, denote $\theta$ by $\theta_{BH}^i$, and for the $i$-th exhibit in training set $C$, denote $\theta$ by $\theta_{HC}^i$, where $H$ denotes HM model.

3.3. Nonparametric models

The AR(1) model and the HM model both assume a Normal distribution for the error terms. A nonparametric model in which this assumption is dispensed with is also used to
This package sets $h_k$ is then used to select the values of $w$ from the point $h$ and a $k$-th nearest neighbour adaptive bandwidth (Breiman et al. 1977). The functions $h_k$ and $w$ are given by

$$h_k = \frac{1}{(n_k^D - 1)h_k} \sum_{j=2}^{j=n_k^D} K_j \left( \frac{z_{i;j} - w_{i,j-1}}{h_k} \right)$$

and

$$w_k = \frac{1}{(n_k^D - 1)h_k} \sum_{j=2}^{j=n_k^D} K_j \left( \frac{z_{i;j} - w_{i,j-1}}{h_k} \right)$$.

Here, $h_1$, $h_2$, and $h_3$ are bandwidths, and $K_1$, $K_2$, and $K_3$ are kernel functions, see Fan et al. (1996), Hall et al. (2004) and Silverman (1986) for further details. An earlier application of kernel density estimation for independent observations in forensic science is given in Aitken and Taroni (2004). The Gaussian kernel

$$K(s) = (2\pi)^{-1/2} \exp(-s^2/2)$$

is used here for all three functions $K_1$, $K_2$, and $K_3$.

The functions $\hat{f}_{iD}$ for each $i \in \{1, 2, \ldots, m_D\}$ can be calculated in R using the np package (Hayfield, 2008). This package sets $h_2 = h_3$ (the bandwidths that apply to the previous banknote in the numerator and denominator, respectively), and finds the optimal bandwidths $h_1$ and $h_2$ using cross-validation and maximising the estimated likelihood (Silverman, 1986). The functions $\hat{f}_{iD}$ are calculated using two different bandwidth types, a fixed bandwidth and a $k$-th nearest neighbour adaptive bandwidth (Breiman et al., 1977). For the adaptive bandwidth, $h$ should be replaced with $h_j$, where $h_j$ is the Euclidean distance from either the point $w_{i,j}$ to the $k_1$-th nearest sample point (if bandwidth $h_1$ is under consideration), or from the point $w_{i,j-1}$ to the $k_2$-th nearest sample point (if bandwidths $h_2$ or $h_3$ are under consideration, since $h_2 = h_3$) (Terrell and Scott, 1992, Wilson et al., 2014). Cross-validation is then used to select the values of $k_1$ and $k_2$ that maximise the estimated likelihood.

4. Classification for a set of notes of unknown type

In this section, methods for the calculation of the general likelihood $f(z \mid H_D)$ are given with the results providing input to the LR $f(z \mid H_C)/f(z \mid H_B)$. 
4.1. Autoregressive model

The likelihood \( f(z | H_D) \) is estimated by

\[
f(z | H_D) = \int f(z_1 | \theta^D_H) f(z_2 | z_1, \theta^D_H) \ldots f(z_n | z_{n-1}, \theta^D_H) f(\theta^D_H) \, d\theta^D_H
\]

\[
\approx \sum_{i=1}^{i=m_D} v_i \int f(z_1 | \theta^D_A) f(z_2 | z_1, \theta^D_A) \ldots f(z_n | z_{n-1}, \theta^D_A) f(\theta^D_A) f(\theta^D_A | w_i) \, d\theta^D_A
\]

(3)

where the weights \( v_i \) are given by \( v_i = n^D_i / \sum_{i=1}^{i=m_D} n^D_i \). Draws from the posterior distributions \( f(\theta^D_A | w_i) \) can be obtained using a Metropolis-Hastings sampler (as described in section 3.1). Using these draws, each integral in (3) can be estimated using Monte Carlo integration, as described in Wilson et al. (2014).

4.2. Hidden Markov model/ Autoregressive model

Denote the hidden Markov model by \( M_H \) and the autoregressive model by \( M_A \). A comparison of the marginal likelihoods \( f(w_i | M_H) \) and \( f(w_i | M_A) \) provides a check of which of the two models better fits the data \( w_i \). The method developed by Chib and Jeliazkov (2001) can be used to calculate these marginal likelihoods when the draws (of \( \theta^D_H \) or \( \theta^D_A \)) from the Metropolis-Hastings sampler given in the Appendix have converged to the required posterior distribution. If the sampler has not converged for \( \theta^D_H \) or \( \theta^D_A \), Monte Carlo integration can be used, so that for \( r \) draws from the priors of either the HM model or autoregressive model, \( \theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(r)} \), the value \((1/r) \sum_{j=1}^{j=r} f(w_i | \theta^{(j)})\) is used as an estimate of the marginal likelihood for that model. The value \( f(w_i | \theta^{(j)}) \) for the HM model can be calculated using the method described in the Appendix. The model with the larger marginal likelihood can then be selected for each of the \( m_B \) samples and \( m_C \) exhibits. Let \( M = (M_1, \ldots M_{m_D}) \) be the vector representing the model choice for the \( m_D \) samples or exhibits in the general set \( D \). For each entry in the vector, \( M_i \), let \( M_i = M_H \) if the HM model has been chosen for the \( i \)-th sample or exhibit and \( M_i = M_A \) if the autoregressive model has been chosen for the \( i \)-th sample or exhibit. If \( M_i = M_H \) then the parameters used in the model for the \( i \)-th sample or exhibit are \( \theta^D_H \), which is the nine-dimensional vector of HM model parameters and if \( M_i = M_A \) then the parameters used are \( \theta^D_A \), which is the four-dimensional vector of autoregressive model parameters.

Let the random variable representing the model choice for the seized sample \( z \) be \( M_z \). The likelihood \( f(z | H_D) \) is approximated by
\[
f(z \mid H_D) = P(M_z = M_H \mid H_D) f(z \mid H_D, M_z = M_H) + P(M_z = M_A \mid H_D) f(z \mid H_D, M_z = M_A)
\]

\[
= P(M_z = M_H \mid H_D) \int f(z \mid \theta_H^0, M_z = M_H) f(\theta_H^0 \mid w) \, d\theta_H^0 \\
+ P(M_z = M_A \mid H_D) \int f(z \mid \theta_A^0, M_z = M_A) f(\theta_A^0 \mid w) \, d\theta_A^0
\]

\[
\approx P(M_z = M_H \mid H_D) \sum_{i:M_i = M_H}^m P(\theta_H^D = \theta_H^0) \int f(z \mid \theta_H^0, M_z = M_H) f(\theta_H^0 \mid w_i) \, d\theta_H^0 \\
+ P(M_z = M_A \mid H_D) \sum_{i:M_i = M_A}^m P(\theta_A^D = \theta_A^0) \int f(z \mid \theta_A^0, M_z = M_A) f(\theta_A^0 \mid w_i) \, d\theta_A^0
\]

\[
\approx \sum_{i:M_i = M_H}^m v_i \int f(z \mid \theta_H^D, M_z = M_H) f(\theta_H^D \mid w_i) \, d\theta_H^D \\
+ \sum_{i:M_i = M_A}^m v_i \int f(z \mid \theta_A^D, M_z = M_A) f(\theta_A^D \mid w_i) \, d\theta_A^D
\] (4)

where \( v_i \) is a weight assigned to the \( i \)-th sample or exhibit, which is used to approximate \( P(M_z = M_H \mid H_D)P(\theta_H^0 = \theta_H^0) \) or \( P(M_z = M_A \mid H_D)P(\theta_A^0 = \theta_A^0) \). Let \( v_i = n_i^D / \sum_{i=1}^{m_D} n_i^D \) so that the weight varies with the number of banknotes in the sample or exhibit.

Draws from the distributions \( f(\theta_H^D \mid w_i) \) for \( i \in \{1, \ldots, m_D\} \) can be obtained using the Metropolis-Hastings sampler for the HM model, described in the Appendix. The likelihood \( f(z \mid H_D) \) can then be estimated using (4), where the individual integrals under the summation are estimated using Monte Carlo integration. Let \( \theta_H^{D(1)}, \theta_H^{D(2)}, \ldots, \theta_H^{D(r)} \) be \( r \) draws from the distribution \( f(\theta_H^D \mid w_i) \). Then the integral \( \int f(z \mid \theta_H^D, M_z = M_H) f(\theta_H^D \mid w_i) \, d\theta_H^D \) is estimated by \( (1/r) \sum_{j=1}^{r} f(z \mid \theta_H^{D(j)}, M_z = M_H) \). The integrals involving the autoregressive model can be calculated similarly, substituting \( \theta_A^D \) for \( \theta_H^D \).

The value of \( f(z \mid \theta_H^D, M_H) \) is obtained using the method described in the Appendix for the calculation of the likelihood when the HM model is assumed.

Care has to be taken with the interpretation of the results because the estimates of the numerator and the denominator of the LR are obtained using Monte Carlo integration and so will have an associated error. For each seized sample, the LR was estimated three times, each time using a randomly drawn set of 5,000 of the 20,000 posterior draws obtained for each sample and exhibit to estimate each integral (so 5,000 draws were used per integral). The ratio of the maximum and the minimum of these three estimates was calculated. If and only if this ratio was larger than two, the LR was re-estimated five times, each time using 20,000 posterior draws to estimate each integral. This repeat calculation was required for 29 of the 263 seized samples tested. The LR was estimated five times to allow assessment of the range of values obtained. The LRs used to present the results are the average of the LRs obtained using the procedure described above (so based on either 15,000 or 100,000 evaluations for each integral).
The calculation of the likelihood ratio for the HM model is computationally expensive. In practice, the procedure described above can be followed to strike a balance between accuracy and computation time. The range of LRs obtained can then be inspected; if this range is too large, the result would have to be deemed inconclusive. To reduce the probability of false support for the prosecution proposition, \( H_C \), the lower end of the range should be the result reported.

### 4.3. Nonparametric models

For the nonparametric model, the likelihood \( f(z \mid H_D) \) is estimated by

\[
f(z \mid H_D) = f(z_1 \mid H_D)f(z_2 \mid z_1, H_D) \ldots f(z_n \mid z_{n-1}, H_D) \\
\approx \sum_{i=1}^{m_D} v_i f_{D_i}(z_1 \mid H_D) f_{D_i}(z_2 \mid z_1, H_D) \ldots f_{D_i}(z_n \mid z_{n-1}, H_D)
\]

where, as before, \( v_i \) is a weight assigned to the \( i \)-th sample or exhibit, taken to be \( v_i = n_i^D / \sum_{i=1}^{m_D} n_i^D \) and \( f_{D_i}(z_i \mid z_{i-1}) \) is the conditional density function associated with the \( i \)-th sample or exhibit, but evaluated for \( z_i \) and \( z_{i-1} \) of the seized sample. \( f_{D_i}(z_i) \) is the marginal density function associated with the \( i \)-th sample or exhibit but evaluated at the point \( z_i \).

The estimates of \( f_{D_i}(\cdot \mid \cdot) \) from (2) should be substituted into (5) to obtain:

\[
\hat{f}(z \mid H_D) \approx \sum_{i=1}^{m_D} v_i \hat{f}_{D_i}(z_1 \mid H_D) \hat{f}_{D_i}(z_2 \mid z_1, H_D) \ldots \hat{f}_{D_i}(z_n \mid z_{n-1}, H_D),
\]

which is used to estimate the likelihood of the proposition, \( H_B \) or \( H_C \), given the data \( z \).

### 5. Results

These results are based on 29 criminal cases (consisting of 70 exhibits) and 193 general circulation samples.

#### 5.1. Parameter estimation

Posterior distributions of all of the parameters for each of the 70 crime exhibits and 193 general circulation samples, were estimated for both the autoregressive model and the HM model. For the 193 general circulation samples, the posterior distribution of the parameters \( \theta_{C_i}^B = (\mu_i^B, (\sigma_i^B)^2, \alpha_i^B, \beta_i^B) \) and \( \theta_{C_i}^H = (\mu_i^B, \mu_i^H, (\sigma_i^H)^2, (\sigma_i^B)^2, \alpha_i^B, \alpha_i^H, \beta_i^B, \beta_i^H) \), conditional on each sample \( x_i \), were estimated. For the 70 crime exhibits, posterior distributions of the parameters \( \theta_{A_i}^C = (\mu_i^C, (\sigma_i^C)^2, \alpha_i^C, \beta_i^C) \) and \( \theta_{A_i}^H = (\mu_i^C, \mu_i^H, (\sigma_i^H)^2, (\sigma_i^C)^2, \alpha_i^C, \alpha_i^H, \beta_i^C, \beta_i^H) \), conditional on each exhibit \( y_i \), were estimated.

The posterior distributions of the autoregressive means of the general circulation samples \( (\mu_i^B) \) occupied a similar range of values to the posterior distributions of the two lower
means, $\mu_1^B$ and $\mu_1^C$, of the HM model (for general circulation samples and crime exhibits respectively) and also to the higher mean of the HM model for general circulation samples ($\mu_2^B$). Although there was a large overlap between these four posterior distributions, and the posterior distributions of $\mu^C$ and $\mu_2^C$ (the autoregressive mean for crime exhibits and the higher mean of the HM model for crime exhibits), in general the posterior distributions of $\mu^C$ and $\mu_2^C$ were to the higher upper end of the posterior distributions of $\mu^B$, $\mu_1^B$, $\mu_1^C$ and $\mu_2^B$. This is as expected: the crime exhibits should have higher quantities of contamination than the general circulation samples, and the higher means within the HM models should be larger than the lower means. Twelve exhibits (out of 70) were declared as contaminated by experts. These exhibits had higher means than most of the other exhibits.

The posterior distributions of the variances of the autoregressive process ($(\sigma^B)^2$ and $(\sigma^C)^2$) were similar to those of the HM model ($(\sigma_1^B)^2$, $(\sigma_2^B)^2$, $(\sigma_1^C)^2$ and $(\sigma_2^C)^2$). This is unsurprising, as the HM model contains a greater number of parameters, which will reduce the variance. The posterior distributions of $(\sigma_1^B)^2$ and $(\sigma_2^B)^2$ had a slightly larger variance than those of $(\sigma_1^C)^2$ and $(\sigma_2^C)^2$, with the same observed for $(\sigma^B)^2$ in comparison to $(\sigma^C)^2$.

The posterior distributions of the autocorrelation parameters varied between samples and exhibits, but the majority generally had ranges between 0.0 and 0.7. The posterior distributions of the transition probabilities generally had ranges between 0.0 and 0.5, with around three quarters of the posterior distributions having a mode around 0.05. Most of the weight of these posterior distributions was on transition probabilities between 0.0 and 0.25. The prior distribution of the transition probabilities was $B(0.6, 4.0)$, truncated to lie between $2/n_y^D$ and $(n_y^D - 2)/n_y^D$. The position of the mode of this distribution varies, depending on $n_y^D$, but for a sample or exhibit with 100 banknotes, the mode is around 0.05, and most of the weight of the distribution is given to values between 0 and 0.4. This prior was chosen as it was thought that banknotes of either high or low contamination would cluster together (i.e. so that the transition probabilities were less than 0.5). The posterior distributions obtained had most of their weight placed on transition probabilities between 0 and 0.25; this suggests that banknotes with similar quantities of contamination are clustered together more than would be implied by the prior.

### 5.2. Likelihood ratios

Rates of misleading evidence were calculated for each of the five models fitted: the HM model, the AR model of order one, the nonparametric model with fixed bandwidth, the nonparametric model with adaptive bandwidth and the standard model, with results given in Table 1. Results for the last four models are also given in Wilson et al. (2014). In order to calculate these rates, each exhibit and sample was taken out in turn and treated as the seized sample of banknotes, $z$. The other samples and the exhibits from other criminal cases (i.e., removing the chosen exhibit or sample and any exhibits from the same criminal case as the chosen exhibit if the seized sample were known to be from the set $C$) were then treated as the measurements $y$ and $x$, respectively, and used to calculate the likelihood ratio for the seized sample. Note that for the AR model, one sample was removed from training set $B$ and for the HM model three exhibits and five samples were removed from the training sets, due to nonconvergence of the Metropolis-Hastings sampler, but these exhibits and samples were still treated as seized samples for the purpose of estimating rates of misleading evidence. If the likelihood ratio for the seized sample were greater than one,
Table 1. Rates of misleading evidence estimated as \((r/n)\) where \(r\) is the number of exhibits / samples out of \(n\) analysed for which the likelihood ratio is said to be misleading in each context.

<table>
<thead>
<tr>
<th></th>
<th>Hidden Markov Model</th>
<th>AR(1) fixed bw</th>
<th>Nonparametric adaptive nn</th>
<th>Standard model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exhibit</td>
<td>0.357 (25/70)</td>
<td>0.371 (26/70)</td>
<td>0.271 (19/70)</td>
<td>0.257 (18/70)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.500 (35/70)</td>
<td></td>
</tr>
<tr>
<td>Sample</td>
<td>0.104 (20/193)</td>
<td>0.155 (30/193)</td>
<td>0.321 (62/193)</td>
<td>0.269 (52/193)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.135 (26/193)</td>
<td></td>
</tr>
</tbody>
</table>

\(z\) was said to provide support for proposition \(H_C\) and if it was less then one, \(z\) was said to provide support for proposition \(H_B\). The proportion of times that the seized sample provided support for the proposition with which it was not associated is given as the rate of misleading evidence.

The two nonparametric models have the smallest rates of misleading evidence for crime exhibits. The method with adaptive bandwidth has a slightly smaller rate of misleading evidence than the method with a fixed bandwidth. These rates are, however, quite large; 25.7% of the evidence provided by crime exhibits is said to be misleading for the nonparametric model with adaptive bandwidth. As discussed in Section 2, some of the crime exhibits are not contaminated any more than general circulation samples, and yet are still included in training dataset \(C\). As a result, low rates of misleading evidence are not expected for crime exhibits, even for well fitting models, and hence these rates should not be used for model comparison.

The rates of misleading evidence for general circulation samples are smaller; the best is achieved with the HM model, which has a rate of misleading evidence of 10.4%. The AR model also performs well, with a rate of 15.5%. The standard model had a rate of misleading evidence of just 13.5% of samples, which is higher than the HM model, and slightly lower than the AR model. The nonparametric models performed badly here, with 32.1% and 26.9% of samples said to be misleading for the fixed bandwidth and adaptive bandwidth models, respectively.

The rates of misleading evidence in Table 1 show which of the models is best at minimising the number of occasions on which a piece of evidence is found to be misleading, but they do not give an indication of the size of the likelihood ratios obtained. A likelihood ratio value close to one does not provide strong support for either proposition. A method which gives larger likelihood ratios, where these large values are justified, is of more value when being used in evidence evaluation. The plots in Figures 3 and 4 consider the actual values of the likelihoods and the likelihood ratios for the models under discussion.

The Tippett plots in Figure 3 indicate the proportion of general circulation samples (dashed) and crime exhibits (solid) that have a log-likelihood ratio greater than the value on the horizontal axis of the plot. Figure 3 shows that the AR model has log-likelihood ratios which are much closer to the neutral value of zero than the HM model and the two nonparametric models, meaning that it is less useful for discriminating between crime exhibits and general circulation samples. The standard model has larger log-likelihood
Fig. 3. Tippett plots of log likelihood ratio values. Clockwise from top left - hidden Markov model, AR(1), adaptive bandwidth, standard model, fixed bandwidth. The last four plots are also given in Wilson et al. (2013) and are reprinted with permission from Elsevier.
Fig. 4. Scatter plots of log likelihood values without outliers. Clockwise from top left - hidden Markov model, AR(1), adaptive bandwidth, standard model, fixed bandwidth, with the line of equality of log likelihood values shown.
ratios than the AR model for crime exhibits that are not misleading, but has log-likelihood ratios similar in size (so close to zero) to those seen for the AR model for general circulation samples.

The two nonparametric models give larger log likelihood ratios in absolute terms than all of the parametric models, but in some cases these large values are misleading because some general circulation samples have very large and positive log-likelihood ratios. The standard model also gives a large and positive misleading log-likelihood ratio to one of the general circulation samples. The HM model and the AR model have small log-likelihood ratios for general circulation samples that provide misleading support, which means that seized samples that are from general circulation are unlikely to provide strongly misleading evidence with these models.

For the two exhibits which, when treated as the seized sample have the largest log-likelihood ratios using the HM model, one had a slightly unstable log-likelihood ratio estimate, with estimates ranging from 16 to 18, and the other had a very unstable log-likelihood ratio estimate, with estimates ranging from 12 to 30. Therefore, the large tail on the Tippett plot for crime exhibits when the HM model is used may be misleading. However, log-likelihood ratio values for other exhibits seem to be slightly larger when the HM model is used, in comparison to the AR model, and log-likelihood ratio values for general circulation samples are much smaller, which suggests that the model is supporting the correct proposition more effectively than the AR model.

Scatter plots for the samples and exhibits plotted according to their log-likelihoods under propositions $H_C$ and $H_B$ showed several outliers relating to the nonparametric models. These outliers are not displayed in Figure 4; the scales of the axes of the plots in Figure 4 were adjusted so that the log-likelihoods of the other samples and exhibits could be seen more clearly. These outliers relate to four crime exhibits that have much larger log-likelihood values under $H_C$ than under $H_B$. When the AR model and HM model are used, these exhibits have log-likelihood values under $H_C$ which are similar to the log-likelihoods under $H_B$. These particular four results occur in exhibits with over 1000 banknotes and may be indicative of a lack of robustness in the nonparametric methods for samples or exhibits with a large number of banknotes. Large samples or exhibits may contain some banknotes with contamination quantities that are very unusual (this is likely, given their size). Then, the estimated conditional density functions $\tilde{f}_D$ will need to be evaluated in their tails for these unusually contaminated banknotes. There are problems (e.g., see Silverman, 1986) associated with the use of kernel density estimates for the estimation of the tails of density functions. These problems could be the cause of the large errors seen for these four exhibits. It is interesting to note that the problems with these four exhibits are seen for both the fixed and variable bandwidth models, so the use of a variable bandwidth does not seem to have resolved the problem for these data.

Scatter plots for the four models accounting for autocorrelation in Figure 4, omitting exhibits or samples with very small likelihoods under either of the propositions, show that the log-likelihood values cluster around the line in the case of the AR model, suggesting equal values under both propositions, and are slightly further from the line for the HM model, and are furthest from the line for the nonparametric models. This clustering around the line corresponds to conclusions drawn from the Tippett plots, that the HM model and the nonparametric models provide the largest absolute values of log-likelihood ratios.

The log-likelihood values for the standard model are much smaller in absolute value than
those of the other models especially noting the difference in the scales of the axes. This is because the means of the measurements for each sample and exhibit are used directly in the calculation of the likelihoods, rather than the individual measurements on each banknote. As the means for each sample and exhibit are used directly in the calculation of the likelihoods, the number of banknotes in each sample or exhibit does not have as big an effect on the log-likelihood values as it does for the other models. As a result, the points do not extend in a diagonal line. The cluster of crime exhibits in the top left corner of the scatter plot for the standard model relates to crime exhibits which have means that are in the right hand mode of the density plot of the means that can be seen in Wilson et al, (2014); i.e., these exhibits have means that are dissimilar to means seen for general circulation.

Some general circulation samples have large log-likelihood values under proposition $H_C$ in comparison to their log-likelihood values under proposition $H_B$ when the nonparametric models are used. These can be seen in Figure 4. It would be hoped that the log-likelihood values under $H_C$ would be smaller than the log-likelihood values under $H_B$ for these samples.

Overall, based on the rates of misleading evidence for the general circulation samples, and the problems with outliers, it is recommended that the parametric models are used to calculate likelihood ratios for these data. The HM model has a slightly smaller rate of misleading evidence than the AR model and the standard model, and discriminates slightly better by returning larger absolute log likelihood ratio values when samples and exhibits are not said to be misleading, making this model more useful in discriminating between general circulation samples and crime exhibits in practice.

The HM model allows for two different levels of contamination within samples and exhibits, as well as accounting for autocorrelation between measurements on adjacent banknotes. By not accounting for autocorrelation, there is a risk that likelihood ratios will be overstated as it is known that autocorrelation is present. By not allowing for different levels of contamination on different bundles within samples and exhibits, there is potential for understating likelihood ratios. However, care should be taken when using the HM model, as the Monte Carlo error can be large when estimating the LR, so (particularly with large seized samples) the estimates of the LR are subject to a greater variance than those of the AR model and the standard model. For example, when the LR was estimated with 20,000 posterior draws for each integral, one exhibit had log likelihood ratio estimates ranging from 16 to 19 and another exhibit had log likelihood ratio estimates ranging from 12 to 30. However, only five other samples or exhibits (out of 263) had a range for the LR (as distinct from log LR) in which the ratio of the maximum to minimum was greater than 10.

6. Conclusion

Previous models used to evaluate LRs for cocaine quantities on banknotes have assumed independence between adjacent banknotes (Besson, 2004). Four models have been developed which do not make this assumption. An autoregressive model of order one, a HM model and two nonparametric models have been developed to model the quantities of cocaine on banknotes from general circulation and on banknotes which are associated with a person who is involved with drug crime involving cocaine. The results are compared with those from a model that assumes independence. The likelihood ratio for $H_C$ and $H_B$ was calculated using these five models. The best performing model of the five was found to be the HM
model. This model had a rate of misleading evidence for general circulation samples of 10.4 
%, and returned larger absolute log-likelihood ratio values than the autoregressive model for 
crime exhibits and general circulation samples that were not found to provide misleading 
evidence when treated as the seized sample.

Previous work in the area of drug traces on banknotes (Besson (2004), Dixon (2006), 
Jourdan et al., (2013)) has defined the database $C$ to be banknotes that have been seized 
by law enforcement agencies. Banknotes in such a database are not necessarily associated 
with a person who is involved with drug crime. This work improves on this definition by 
establishing a database of banknotes for $C$, where the suspect from whom the banknotes 
were seized has subsequently been convicted of a crime involving cocaine. There is still work 
to be done, as the evidence of the drug traces on the banknotes was, in some cases, used as 
evidence to convict the suspect. Ideally a database of banknotes which were known to have 
been contaminated in the course of illegal activity involving cocaine would be constructed, 
without evidence relating to the quantities of cocaine on those banknotes influencing the 
trial, although collection of such data may be impractical.

The models described here use univariate data, the logarithm of the peak area of the 
cocaine product ion $m/z$ 105. Data on the peak maximum for this cocaine product ion, and 
also both the peak area and peak maximum for four other drugs are available. Future work 
will generalise the univariate approach and consider this multivariate problem.

Further details of a forensic science nature may be found in an associated paper, Wilson 
et al. (2014).

Acknowledgements
The research was supported by an EPSRC CASE award, voucher number 009002219. We 
are grateful to the Editor and a referee for very helpful comments on an earlier version of 
the paper.

References
Scientists, second edition. Chichester: John Wiley and Sons Ltd.

time series subject to Markov mean and variance shifts. Journal of Business and Economic 

sanne.

Breiman,L., Meisel,W. and Purcell,E. (1977) Variable kernel estimates of multivariate 
densities. Technometrics, 19(2), 135-144.


don: Chapman and Hall.


Richardson, S. and Green, P. J. (1997) On Bayesian Analysis of Mixtures with an Un-
The Evaluation of Evidence for Autocorrelated Data in Relation to Traces of Cocaine on Banknotes

known Number of Components (with discussion), *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(4), 731-792


7. Appendix

7.1. Hidden Markov Model

Calculation of the likelihood

Consider a general sample of data \( w_i = \{w_{it}, t = 1, \ldots, n_i^D\} \), where \( w_i \) should be replaced by \( x_i \) (a sample of general circulation banknotes), \( y_i \) (an exhibit of criminal case banknotes) or \( z \) (a seized set of banknotes) as required. The general sample of data is associated with a set of hidden states \( S = \{S_t, t = 1, \ldots, n_i^D\} \). The method for calculating the likelihood \( f(w_i | \theta) \), when the HM model is assumed, is outlined in this section. As the HM model is assumed, the parameters are given by \( \theta = (\mu_1, \mu_2, \sigma^2_1, \sigma^2_2, \alpha, \rho_{01}, \rho_{10}, \beta_1, \beta_2) \). As the observations \( w_i \) depend on the parameters \( \theta \) via the hidden states, \( S \), these hidden states must be summed out. The forward-algorithm, as described in Rabiner (1989), with a straightforward adjustment to allow for the dependence of \( w_{it} \) on \( w_{i,t-1} \), is used. The likelihood is given by:

\[
f(w_i | \theta) = \sum_S f(w_i, S | \theta).
\]

As described in Rabiner (1989), this is calculated by recursively calculating the forward variables \( \beta_1(s) \), given by:

\[
\beta_1(s) = f(w_{it} | \theta, w_{i,t-1}, S_t = s) \sum_{k=1}^{4} f(S_t = s | \theta, S_{t-1} = k) \beta_{t-1}(k).
\]
Then, the likelihood is given by:

\[
f(w \mid \theta) = \sum_{s=1}^{4} \beta_{s_i}^{(s)}(s).
\]

In (6), the term \( f(w_{it} \mid \theta, w_{i,t-1}, S_t = s) \) is Normally distributed, with mean \( \mu_{s}^{(1)} - \alpha(w_{i,t-1} - \mu_{s}^{(2)}) \) and variance \( \sigma_s^2 \). These parameters depend on the value of the hidden state \( S_t \). The term \( f(S_t = s \mid \theta, S_{t-1} = k) \) is given by the transition matrix of the hidden states.

This method of calculating the likelihood is used in the Metropolis-Hastings sampler, to calculate the marginal likelihoods \( f(w_i \mid M_H) \), with \( M_H \) denoting the choice of the HM model, and to calculate the likelihoods \( f(z_i \mid H_D) \) where \( D \) is one of \( C \) or \( B \).

**Metropolis-Hastings sampler**

The methods for obtaining draws from the posterior distribution of \( \theta \), conditional on a single sample or exhibit when the HM model is used are described here.

Consider a general set of data \( w_i = \{w_{it}, t = 1, \ldots, n_i^D\} \) with hidden states \( S \). The Metropolis-Hastings sampler is used to obtain draws from the posterior density function \( f(\theta \mid w_i) \).

The posterior distribution of the hidden states is not required to calculate the LR of a seized sample, so to improve mixing, the hidden states are summed out of the likelihood, \( f(w_i, S \mid \theta) \), to obtain the required likelihood \( f(w_i \mid \theta) \). This is done using the method in Rabiner (1989), described above.

The parameters \( \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \alpha, p_{01}, p_{10}, \beta_1, \beta_2 \) are given the following prior distributions:

\[
\begin{align*}
\mu_1, \mu_2 & \sim N(\mu_0, V_\mu), \\
\sigma_1^2 & \sim IG(\gamma_1, \beta_1), \\
\sigma_2^2 & \sim IG(\gamma_2, \beta_2), \\
\alpha & \sim N(\alpha_0, V_\alpha)I(|\alpha| < 1), \\
p_{01} & \sim B(a, b)I(2/n_i^D < p_{01} < (n_i^D - 2)/n_i^D), \\
p_{10} & \sim B(a, b)I(2/n_i^D < p_{10} < (n_i^D - 2)/n_i^D), \\
\beta_1, \beta_2 & \sim \Gamma(g, h),
\end{align*}
\]

where \( \mu_0, V_\mu, \gamma_0, V_\alpha, a, b, g \) and \( h \) take the values given in the body of the text. The gamma distribution \( \Gamma(g, h) \) is such that if \( \beta \sim \Gamma(g, h) \) then

\[
f(\beta \mid g, h) = \frac{h^g}{\Gamma(g)} \beta^{g-1} e^{-h \beta}; \ g > 0, h > 0, \beta > 0,
\]

where \( \Gamma(.) \) is the gamma function. The inverse gamma distribution \( IG(\gamma, \beta) \) is such that if \( \sigma^2 \sim IG(\gamma, \beta) \) then
\[
f(\sigma^2 \mid \gamma, \beta) = \frac{\beta \gamma}{\Gamma(\gamma)} (\sigma^2)^{-(\gamma+1)} e^{-\beta/\sigma^2}; \quad \beta > 0, \gamma > 0, \sigma > 0.
\]

The joint prior distribution, \( f(\theta) \), is therefore given by:

\[
f(\theta) = f(\mu_1) f(\mu_2) f(\sigma_1^2) f(\sigma_2^2) f(\beta_1) f(\beta_2) f(\alpha) f(p_{01}) f(p_{10})
\]

\[
\propto \exp \left[-\frac{1}{2V_\mu} (\mu_1 - \mu_0)^2 \right] \times \exp \left[-\frac{1}{2V_\mu} (\mu_2 - \mu_0)^2 \right]
\]

\[
\times \beta_1^{2\gamma} \sigma_1^{-(2\gamma+2)} \exp \left[-\frac{1}{\sigma_1^2} \right] \beta_1^{\gamma-1} \exp(-h\beta_1)
\]

\[
\times \beta_2^{2\gamma} \sigma_2^{-(2\gamma+2)} \exp \left[-\frac{1}{\sigma_2^2} \right] \beta_2^{\gamma-1} \exp(-h\beta_2)
\]

\[
\times \exp \left[-\frac{1}{2V_\alpha} (\alpha - \alpha_0)^2 \right] I(|\alpha| < 1)
\]

\[
\times p_{01}^{\alpha-1} (1 - p_{01})^{1-\alpha} p_{10}^{\beta-1} (1 - p_{10})^{1-\beta} I(2/n_i^D < p_{01} < (n_i^D - 2)/n_i^D) I(2/n_i^D < p_{10} < (n_i^D - 2)/n_i^D).
\]

Denote the parameters at step \( j \) of the Metropolis-Hastings algorithm by \( \theta^{(j)} = (\mu_1^{(j)}, \mu_2^{(j)}, \sigma_1^{2(j)}, \sigma_2^{2(j)}, \alpha^{(j)}, p_{01}^{(j)}, p_{10}^{(j)}, \beta_1^{(j)}, \beta_2^{(j)}) \) and the proposed parameters as \( \theta' = (\mu_1', \mu_2', \sigma_1'^2, \sigma_2'^2, \alpha', p_{01}', p_{10}', \beta_1', \beta_2') \), then the Metropolis-Hastings sampler updates the parameters as follows:

\[
\mu_1' = \mu_1^{(j)} + \epsilon_1,
\]

\[
\mu_2' = \mu_2^{(j)} + \epsilon_2,
\]

\[
\log(\sigma_1'^2) = \log(\sigma_1^{2(j)}) + \epsilon_3,
\]

\[
\log(\sigma_2'^2) = \log(\sigma_2^{2(j)}) + \epsilon_4,
\]

\[
\alpha' = \alpha^{(j)} + \epsilon_5,
\]

\[
\log(p_{01}'/(1 - p_{01}^{(j+1)})) = \log(p_{01}^{(j)}/(1 - p_{01}^{(j)})) + \epsilon_6,
\]

\[
\log(p_{10}'/(1 - p_{10}^{(j+1)})) = \log(p_{10}^{(j)}/(1 - p_{10}^{(j)})) + \epsilon_7,
\]

\[
\log(\beta_1') = \log(\beta_1^{(j)}) + \epsilon_8,
\]

\[
\log(\beta_2') = \log(\beta_2^{(j)}) + \epsilon_9.
\]

Here, \( \epsilon_k \) is a Normally distributed random variable, with zero mean and variance \( V_k \) for \( k \in \{1, 2, \ldots, 9\} \). It has been shown (Gelman et al., 1996) that the \( V_k \) should be chosen so that the proportion of accepted updates is close to 25%.

The updated parameter \( \theta' \) is accepted (and \( \theta^{(j+1)} \) is set to \( \theta' \)) if \( U < \min(1, A) \), where \( U \) is drawn from a uniform distribution on the interval \([0, 1]\), and \( A \) is given by:

\[
A = \frac{f(\mathbf{w}_1 \mid \theta')} {f(\mathbf{w}_1 \mid \theta^{(j)})} \frac{f(\theta') \sigma_1^{2(j)} \sigma_2^{2(j)} \beta_1^{(j)} \beta_2^{(j)} (1 - p_{01}^{(j)}) p_{01}^{(j)} (1 - p_{10}^{(j)}) p_{10}^{(j)}} {f(\mathbf{w}_1 \mid \theta^{(j)}) f(\theta^{(j)}) \sigma_1^{2(j)} \sigma_2^{2(j)} \beta_1^{(j)} \beta_2^{(j)} (1 - p_{01}^{(j)}) p_{01}^{(j)} (1 - p_{10}^{(j)}) p_{10}^{(j)}}.
\]
If \( \theta' \) is not accepted then \( \theta^{(j+1)} \) is set to \( \theta^{(j)} \). The term \( \sigma_1'^2 \sigma_2'^2 \beta_1'(1-p_0')p_{01}(1-p_{10}')p_{10}' \) in the numerator of (7) and the term \( \sigma_1^{2(j)} \sigma_2^{2(j)} \beta_1^{(j)} \beta_2^{(j)} (1-p_{01}')(1-p_{10}')(1-p_{01}')(1-p_{10}') \) in the denominator are the Jacobians of the log transformations of \( \sigma_1, \sigma_2, \beta_1 \) and \( \beta_2 \) and the logistic transformations of \( p_{01} \) and \( p_{10} \).

As discussed in Scott (2002) and Frühwirth-Schnatter (2001), the likelihood of \( \theta \) is invariant under certain permutations of the states. In particular, the likelihood is unchanged if the state labels (1,2,3,4) are swapped with (4,3,2,1). As such, the marginal posterior distributions of each of the parameters should be bimodal (with the exception of \( \alpha \), which is not state dependent). Frühwirth-Schnatter (2001) suggests using a permutation sampler to improve the mixing of a Markov chain Monte Carlo sampler where this label switching problem exists. At the end of each run of the sampler, the labels are permuted from (1,2,3,4) to (4,3,2,1) with probability 0.5. This is added as a final step to the algorithm as follows:

- Obtain a sample \( b \) from a Bernoulli random variable with parameter \( p = 0.5 \).
- If \( b = 0 \), take no action. If \( b = 1 \), swap \( \mu_{1}^{(j+1)} \) with \( \mu_{2}^{(j+1)} \), \( p_{01}^{(j+1)} \) with \( p_{10}^{(j+1)} \), \( \beta_{1}^{(j+1)} \) with \( \beta_{2}^{(j+1)} \) and \( \sigma_{1}^{2(j+1)} \) with \( \sigma_{2}^{2(j+1)} \).

Draws from the posterior distribution of \( \theta \) are used in evaluations of the probability density function of measurements on a seized sample. As this too will be invariant under permutation of the state labels, it is not necessary to constrain the sampler by insisting that \( \mu_1 < \mu_2 \). Instead, draws from the bimodal posterior distributions can be used directly in the evaluations of the probability density function of the seized sample.