The Annals of Statistics
1991, Vol. 19, No. 4, 2067–2089

BELIEF TRANSFORMS AND THE COMPARISON OF HYPOTHESES

BY MICHAEL GOLDSTEIN

University of Durham

We describe a general approach to the comparison of two stochastic specifications over a collection of random quantities and then extend the comparison to collections of stochastic specifications. This comparison derives from the eigenstructure of the belief transform, which we construct in full generality for partially specified belief structures. We describe an application of the methodology, namely the comparison of hypotheses. Given various competing probabilistic specifications for a collection of observable quantities, we use the belief transform to separate those quantities for which the different models make very different predictions (and therefore which may be used to distinguish between the models) from those quantities for which the different models make similar predictions (and therefore which may be used to assess the suitability of the general class of models under consideration). Finally, we describe a particular application of hypothesis comparison which relates to the modelling of a collection of time series derived from an aluminium smelting process.

Introduction. Many stochastic analyses require the comparison of two, or several, stochastic specifications over a collection of random quantities. Typical examples of such comparisons are: hypothesis testing (where the comparison is between various competing stochastic models for a collection of observable quantities); experimental design (where the comparison is between stochastic measures of the amount of information to be gained about various observable quantities from the competing designs); and sensitivity analysis (where the comparison is between a given probabilistic specification and various alternate specifications, which are obtained by small perturbations of the original specification).

In this paper, we propose a systematic approach to the comparison of any two specifications and then extend the comparison to collections of specifications. The comparison derives from the eigenstructure of the belief transform, which is the fundamental object of study in this paper. We describe the construction of the belief transform within a subjectivist framework, but the methodology can be similarly applied to compare any stochastic specifications. We build the belief transform over general inner product spaces to allow the possibility of comparing partial belief specifications.

In Section 1, we explain how belief transforms can be constructed in full generality and describe various properties of the transforms.

Received February 1989; revised December 1990.

AMS 1980 *subject classifications*. Primary 62A15; secondary 62G10.

Key words and phrases. Autocorrelated errors, mean comparisons, partial specification, previsision, trend test.

In Section 2, we describe a direct application of the belief transform methodology, which we term the comparison of hypotheses. The idea is that, given various competing (partial) probabilistic specifications for a collection of observable quantities, we use the belief transform to separate those quantities for which the different models make very different predictions (and which may thus be used to distinguish between the models) from those quantities for which the different models make roughly similar predictions (and which may thus be used to assess general acceptability of the class of models under comparison).

In Section 3, we describe a particular application of the notion of hypothesis comparison, which relates to the modelling of a collection of time series derived from an aluminium smelting process.

1. Belief transforms.

1.1. *Belief structures.* In our development, we take as primitive the notion of the prevision of a random quantity. [The prevision, $P(X)$, for a random quantity, X, can be considered to be your (subjectively assessed) expectation for X, but specified directly as a primitive quantity. The probability of an event, H, is identified with the prevision for the indicator function for H; see de Finetti (1974) for a detailed explanation and development of this approach.] We make prevision primitive so that we can analyse restricted aspects of belief specifications, where necessary, without having to first specify detailed probability measures over some limiting partition.

The collection of prevision statements is organised into a belief structure defined as follows. First, specify a base, $C = [X_0, X_1, X_2, \dots]$, of random quantities, including $X_0 = 1$, the unit constant. Construct the vector space, L, whose vectors are the finite linear combinations $\{a_{i_1}X_{i_1} + \cdots + a_{i_k}X_{i_k}\}$ of the elements of C. Thus L is the space of all quantities whose prevision is uniquely determined by specification of previsions for all of the elements of C.

Next specify, for each pair i, j, the prevision $P(X_iX_j)$. The belief structure B denotes the (closure of the) inner product space whose vectors are the elements of L, with inner product $(X, Y) = P(XY)$.

For example, a discrete probability space is represented as a belief structure whose base consists of the indicator functions for the elementary events in the space, so that the vectors correspond to the collection of random variables over the space. Similarly, the usual Bayesian formulation, in terms of a prior measure over a probability space, is represented as a belief structure by the corresponding Hilbert space of square integrable functions over the probability space with respect to the prior measure.

One of the main objectives of the belief structure construction is to simplify belief specification by allowing you to restrict specification to whatever (often small) subspace of the full probabilistic structure is adequate for the problem.

The belief structure construction was used in Goldstein (1981) and formally described in Goldstein (1986), with emphasis on exchangeability between

structures. A general overview of the role of such structures in subjectivist theory is given in Goldstein (1987a, b). Basic operations for manipulating such structures were given in Goldstein (1988a, b). A general computer language for analysing belief structures has been written under a SERC grant. The program is termed [B/D] (an acronym for beliefs adjusted by data) and is described in two technical reports, Goldstein (1987c) and Wooff (1987).

While the above-referenced papers are all phrased in terms of the particular inner product $(X, Y) = P(XY)$, it is important to note that the analysis relates equivalently to any inner product over the chosen base. This paper concerns the comparison of different inner products over the same linear space L, so that we will informally refer to a "generalised" belief structure as any inner product space over L, for which the inner product characterises some aspect of your beliefs over C.

1.2. *Belief transforms.*

1.2.1. *Comparing quadratic forms.* At many stages of a stochastic analysis, you must compare various stochastic specifications. For example:

(A1) You may wish to compare alternative plausible prior specifications, (i) to assess the differences between your beliefs over certain observable quantities which follow from various plausible parametrisations or (ii) to assess the stability of your inferences under small perturbations of your belief inputs or (iii) to compare your current collection of partial prior specifications with various larger collections of specifications, for example, generated by various standard conjugate prior forms, to identify possible gains from increasing your actual number of specifications;

(A2) You may receive new information, in which case you may wish to summarise the important differences between your prior and posterior beliefs, relate these differences to different aspects of the information and identify "influential" aspects of the information;

(A3) You may intend to make certain observations and for planning purposes you may want to identify the nature and magnitude of your expected changes in your beliefs, for example, to compare various experimental designs or sampling frames.

In each case, within our formulation, there will be two or more inner products defined over the same linear space, and we need informative summaries as to the main differences between the inner products. More generally, we may compare the original inner product with a derived quadratic form for which some of the elements have zero norm. (For example, we might compare prior to posterior variances over a linear space, with some posterior variances 0.) Thus, in general, we will compare the initial inner product with various symmetric positive semidefinite sesquilinear functionals.

A symmetric psd sesquilinear functional on B is any real-valued functional, f, on $B \times B$ satisfying, for all X, Y, Z in B, and numbers a, b:

(i) $f(X, Y) = f(Y, X)$,
(ii) $f(X, X) \geq 0$,
(iii) $f(aX + bZ, Y) = af(X, Y) + bf(Z, Y)$.

[Sometimes, such a functional is termed positive symmetric, while if inequality (ii) is strict for X nonzero, f is termed strictly positive.]

A sesquilinear functional is bounded if the infimum of the values of k for which $k\|X\| \, \|Y\| \geq |f(X, Y)|$, for all X, Y in B is finite. This infimum is termed the norm of the functional. [The notation and results on inner product spaces are given in Bachman and Narici (1966).]

In our formulation, we begin with a belief structure B, with inner product $(\cdot, \cdot)$. We compute a bounded symmetric psd sesquilinear functional $\{\cdot, \cdot\}$ over B. Next, we seek ways to summarise the similarities and differences between $\{\cdot, \cdot\}$ and the inner product $(\cdot, \cdot)$. To this end, we introduce the belief transform.

1.2.2. *Defining the belief transform.* We will exploit the following property of inner product spaces.

THEOREM. *A necessary and sufficient condition for f to be a bounded, symmetric positive semidefinite, sesquilinear functional over the Hilbert space B is that f is of the form*

$$f(X, Y) = (X, TY),$$

where T is a bounded self-adjoint operator over B, with norm equal to the norm of f [see Bachman and Narici (1966), Section 21.1].

Thus bounded sesquilinear functionals are equivalent to self-adjoint operators. We refer to this construction as follows.

DEFINITION. The bounded self-adjoint operator T defined by the relations, for all X, Y, that

$$\{X, Y\} = (X, T(Y))$$

is termed the *belief transform* for $(\cdot, \cdot)$ associated with $\{\cdot, \cdot\}$.

One particular type of belief transform, linked to the inner product $(\cdot, \cdot) = P(X - P(X))(Y - P(Y))$ and constructed by a system of orthogonal projection, was analysed in Goldstein (1981). This transform compared $(\cdot, \cdot)$ with $\{X, Y\} = P(X - P_1(X))(Y - P_1(Y))$, where P_1 is the posterior prevision at time 1. This transform analyses the expected changes in belief caused by new information. An overview of the role of this special type of transform is given in Goldstein (1987a, b).

1.2.3. *Interpreting the eigenstructure of the belief transform.* The eigenstructure of the belief transform summarises the differences between the two inner products as follows.

Consider first the case where B is finite dimensional, that is, where the base of B has a finite number of nondependent elements, r say, plus the unit constant. Any symmetric positive semidefinite, sesquilinear functional will be bounded automatically. The associated belief transform T over B will have $r + 1$ orthonormal eigenvectors, $Z_0, Z_1, \ldots, Z_r$, corresponding to eigenvalues $\lambda_r \geq \cdots \geq \lambda_1 \geq \lambda_0 \geq 0$.

Each X in B can be written as

$$X = (X, Z_0)Z_0 + (X, Z_1)Z_1 + \cdots + (X, Z_r)Z_r,$$

so that the ratio $D(X)$ of inner products is given by

$$D(X) = \{X, X\}/(X, X) = \sum \lambda_i (X, Z_i)^2 \Big/ \sum (X, Z_i)^2.$$

Thus $D(X)$ will be large/small if and only if X has large components corresponding to eigenvectors with large/small eigenvalues. In particular, the smallest value of $D(X)$ over all elements of B which are orthogonal to $Z_0, Z_1, \ldots, Z_s$ is λ_{s+1} corresponding to Z_{s+1}, and the largest value of $D(X)$ over all elements orthogonal to $Z_r, Z_{r-1}, \ldots, Z_s$ is λ_{s-1} corresponding to Z_{s-1}. In particular, the norm of T is λ_r.

Thus the eigenvectors lay out a coordinate grid over B which summarises the nature and degree of difference between $(\cdot, \cdot)$ and $\{\cdot, \cdot\}$.

NOTATION. The eigenvectors of the belief transform with eigenvalue not equal to 1 are termed *informative*. Eigenvectors with eigenvalue 1 are termed *noninformative*. The space spanned by the informative eigenvectors is called the *sufficient* subspace for the transform.

NOTE 1. We can define the inverse belief transform, T^{-1}, between $\{\cdot, \cdot\}$ and $(\cdot, \cdot)$ in a similar fashion, that is,

$$(X, Y) = \{X, T^{-1}Y\},$$

provided that we restrict the transform to the strictly positive part of $\{\cdot, \cdot\}$. Note, in particular, that if Z is an eigenvector of T with strictly positive eigenvalue λ, then Z is also an eigenvector of T^{-1}, with eigenvalue $1/\lambda$. Note also that the eigenvectors of T are orthogonal under both $(\cdot, \cdot)$ and $\{\cdot, \cdot\}$.

NOTE 2. The implications of the orthogonality conditions on the eigenvectors depend on the nature of $(\cdot, \cdot)$. In the special case where $(\cdot, \cdot) = P(XY)$, if X_0 is an eigenvector of T, then the orthogonality constraints imply that all of the other eigenvectors have zero prevision and are uncorrelated.

Two types of comparison where this will occur are the following:

1. If we compare quadratic forms which assign the same previsions over the base, then $\{\cdot, \cdot\}$ will typically satisfy, for all Y in B, the relations $\{X_0, Y\} = (X_0, Y)$. Here X_0 is an eigenvector of T, with eigenvalue 1.
2. If we extract X_0 from the system, for example, as part of a linear fitting system, so that $\{\cdot, \cdot\}$ satisfies $\{X_0, X_0\} = 0$, then X_0 is an eigenvector of T, with eigenvalue 0.

NOTE 3. As the eigenstructure of T is determined by the linear space, the decomposition into informative/uninformative eigenvectors is unchanged if we "reparametrise," replacing the base B by any linearly equivalent base of random quantities.

1.2.4. *The complementary belief transform.* It is often useful to work with the complementary belief transform defined as follows.

DEFINITION. The *complementary belief transform* S is defined by

$$S = I - T,$$

where I is the identity operator over B.

The eigenstructures of S and T are essentially equivalent; the eigenvectors are the same, and the eigenvalues of S are 1 minus the corresponding eigenvalues for T. S is psd if and only if the norm of T (i.e., λ_r) is not greater than 1. The transforms S and T decompose the inner product as

$$\begin{aligned}(X, Y) &= \{X, Y\} + ((X, Y) - \{X, Y\}) \\ &= (X, T(Y)) + (X, S(Y)),\end{aligned}$$

where the first term on the right represents the new quadratic form and the second represents the difference between the forms.

In our approach, finite-dimensional spaces are fundamental, in that they represent actual rather than idealised belief specifications. However, if B is not finite dimensional, then T still summarises the relation between the two forms and the spectral representation through the resolution of the identity will provide analogous information to the eigenstructure of T in the finite case. In particular, if the complementary belief transform, S, is compact, then there will be a countable number of eigenvalues of S, tending to 0, and our interpretation for the eigenstructure of T will be as in Section 1.2.3 [see the discussion in Goldstein (1981)].

1.3. *Adjusted belief transforms.* Belief transforms summarise the differences between inner products. Often, it will be useful to make such comparisons in stages. For example, you might wish to analyse the effects of introducing additional parameters upon your beliefs about certain observable

quantities. In such cases it may be natural to introduce the additional parameters one at a time, in order to separate out the overall differences in the final specification into parts attributable to the individual parameters. Alternatively, you might have revised an inner product sequentially, as fresh data came on stream, and now wish to identify the stages of the revision at which the various important changes were made and also whether such changes were gradual and consistent or sharp and contradictory.

Suppose that there are two new inner products $\{\cdot,\cdot\}_1, \{\cdot,\cdot\}_2$ defined over B, with associated belief transforms T_1, T_2 with respect to $(\cdot,\cdot)$.

NOTATION. The belief transform T_{12}, associated with $\{\cdot,\cdot\}_2$, *adjusted* for $\{\cdot,\cdot\}_1$ is the belief transform associated with $\{\cdot,\cdot\}_2$, with respect to the inner product $\{\cdot,\cdot\}_1$, that is, satisfying, for each X, Y in B,

$$\{X, Y\}_2 = \{X, T_{12}(Y)\}_1$$

(provided that the adjusted transform exists, i.e., that $\{\cdot,\cdot\}_2$ is bounded with respect to $\{\cdot,\cdot\}_1$).

Adjusted belief transforms are belief transforms in their own right. We use the adjustment notation in order to relate the inner product $\{\cdot,\cdot\}_2$ to the primary inner product $(\cdot,\cdot)$ through the intermediary product $\{\cdot,\cdot\}_1$. This is analogous to the decomposition of chains of events via probabilistic conditioning, as $P(A \cap B) = P(A|B)P(B)$. For adjusted belief transforms, the decomposition is as follows.

THEOREM 1. *With notation as above,*

$$T_2 = T_1 T_{12}$$

(*operator multiplication is the composition of the two operators*).

PROOF. For each X, Y in B,

$$(X, T_2(Y)) = \{X, Y\}_2 = \{X, T_{12}(Y)\}_1 = (X, T_1(T_{12}(Y))),$$

and the result follows. □

(The same definitions and proof follow when $\{\cdot,\cdot\}_1$ and $\{\cdot,\cdot\}_2$ are positive but not strictly positive. In this case, the requirement that $\{\cdot,\cdot\}_2$ is bounded with respect to $\{\cdot,\cdot\}_1$ implies that the zero elements for $\{\cdot,\cdot\}_1$ are contained in the zero elements of $\{\cdot,\cdot\}_2$.)

The eigenstructure of the adjusted transform T_{12} is interpreted in similar fashion to the eigenstructure of T_2. Large/small eigenvalues identify eigenvectors for which $\{\cdot,\cdot\}_2$ is larger/smaller than $\{\cdot,\cdot\}_1$, and these values are related to the changes between $\{\cdot,\cdot\}_2$ and $(\cdot,\cdot)$ by the general relation given in the theorem.

We have observed that the informative eigenvectors (i.e., with eigenvalue not equal to 1) distinguish the effects of two inner products. From the above theorem, we have the following corollary.

COROLLARY. *Suppose that we have a collection of inner products* $(\cdot, \cdot)_i$, $i = 1, \ldots, k$, *over the linear space* L. *Suppose that there is a linear subspace,* M, *of* L, *with the property that the sufficient subspace for each transform of the form* $T_{i(i-1)}$, $i = 1, \ldots, k-1$, *is contained in* M. *Then the sufficient subspace for each* T_{ji} *is contained in* M.

1.4. *Evaluating the belief transform.*

1.4.1. *Finite-dimensional transforms.* We now give a simple computational algorithm for the belief transform in the special case where L, the underlying linear space, is finite dimensional.

1. Choose a maximal set of vectors, $Z = \{Z_1, \ldots, Z_r\}$, over which $(\cdot, \cdot)$ is strictly positive.
2. Construct the matrix representation, V, of the inner product, $(\cdot, \cdot)$, with respect to Z, that is, $V = (v_{ij})$, an $r \times r$ matrix with $v_{ij} = (Z_i, Z_j)$.
3. Construct the matrix representation, U, of the quadratic form $\{\cdot, \cdot\}$ with respect to Z, that is, $U = (u_{ij})$, $u_{ij} = \{Z_i, Z_j\}$.
4. Evaluate the matrix $W = V^{-1}U$.

W is the matrix representation of T with respect to basis Z.

PROOF. Take any elements X, Y in B. Write X, Y in the coordinate system of Z, that is, write X as $(x_1, \ldots, x_k)$, and Y as $(y_1, \ldots, y_k)$, where

$$X = x_1 Z_1 + \cdots + x_k Z_k, \qquad Y = y_1 Z_1 + \cdots + y_k Z_k.$$

We have $\{X, Y\} = X^T U Y = X^T V V^{-1} U Y = X^T V W Y = (X, TY)$, as required. □

1.4.2. *Adjusted transforms.* Suppose that we construct the two transforms T_1 and T_{12} as in Section 3. Again suppose that L is finite dimensional. For simplicity, suppose also that $\{\cdot, \cdot\}_1$ is strictly positive.

Let V, U_1, U_2 denote the matrix representations of $(\cdot, \cdot), \{\cdot, \cdot\}_1, \{\cdot, \cdot\}_2$, with respect to Z. Then we have as in Section 1.4.1 that, with respect to Z, the matrix representation of T_1 is $V^{-1}U_1$, the matrix representation of T_{12} is $U_1^{-1}U_2$ and the matrix representation of T_2 is $V^{-1}U_2$. (Note that $V^{-1}U_2 = \{V^{-1}U_1\}\{U_1^{-1}U_2\}$, i.e., $T_2 = T_1 T_{12}$ as required.)

(If U_1 is not invertible, then we can substitute a generalised inverse in the above expression, as in our formulation, the zero elements for $\{\cdot, \cdot\}_1$ are contained in the zero elements of $\{\cdot, \cdot\}_2$.)

1.4.3. *Transforms over infinite spaces.* When L is infinite dimensional (as will be the case, for example, when we are analysing a probability model over a continuous space via the representation as the corresponding Hilbert space of

square integrable functions over the space), the analogue to the above matrix representation for the belief transform for $(\cdot, \cdot)$ induced by $\{\cdot, \cdot\}$ replaces the matrix form by the corresponding projection operators.

We create the inner product space A, whose vectors are ordered pairs $[U, V]$, where U, V are elements of L.

The inner product $\langle \cdot, \cdot \rangle$ over A is defined by the following conditions, for any U, V in L:

$$\langle [U,0],[V,0] \rangle = (U,V),$$
$$\langle [0,U],[0,V] \rangle = \{U,V\},$$
$$\langle [U,0],[0,V] \rangle = \{U,V\}.$$

In particular, the subspace B^+ of A consisting of all vectors of form $[U, 0]$ is naturally isometric with the original inner product space, B, with product $(\cdot, \cdot)$, and the subspace B^- consisting of all vectors $[0, V]$ is naturally isometric with the inner product space under $\{\cdot, \cdot\}$.

The orthogonal projection from B^+ to B^- is the transform that takes $[V, 0]$ to $[0, V]$. (Call this transform P^-.) Denote the orthogonal projection from B^- to B^+ by P^+.

THEOREM. *The belief transform T on B is isometrically equivalent to the operator P^+P^- on B^+.*

PROOF. For any U, V in L, we have

$$\begin{aligned}\{U,V\} &= \langle [U,0],[0,V] \rangle = \langle [U,0], P^-[V,0] \rangle \\ &= \langle [U,0], P^+P^-[V,0] \rangle = (U,TV).\end{aligned}$$

□

2. Hypothesis comparison. In this section, we describe an application which shows the belief transform in a particularly simple form.

2.1. *Comparing two simple belief specifications.* Consider the simplest type of belief comparison that we might make. There are two possible collections of beliefs that we may hold concerning a certain collection of observable quantities. If each possible belief specification is in terms of a full joint probability measure, then we may use the methodology of classical or Bayesian hypothesis testing to enable observations on the quantities to help us to distinguish the two possibilities. However, we need methodology to cover situations in which at least one of the following applies:

(i) beliefs are only partially specified;

(ii) we wish to identify which aspects of the two different specifications are responsible for the data appearing to support one or the other hypothesis;

(iii) we wish to reserve the possibility that neither hypothesis applies;

(iv) we would like to distinguish between the hypotheses by a simple and robust analysis, which may be convincing to people holding a wide range of beliefs.

More generally, we may simply be interested in qualitative understanding of the ways in which the hypotheses differ. A simple way to compare the two hypotheses, with respect to some partial collection of belief specifications, is to compare the predictions made by each specification and to identify certain aspects of the data for which the two given belief specifications predict different values. (For example, many simple hypothesis tests are based on the magnitude of some sample average which is expected to be large under one hypothesis and small under the other hypothesis.)

We may similarly identify aspects of the data for which both hypotheses predict roughly comparable values, which are thus primarily diagnostic as to whether either hypothesis can be reasonably supported by the data.

We now describe a systematic method for making such comparisons using the belief transform.

2.2. *Comparison through the belief transform.* Suppose that we are about to observe the values of a collection of random quantities $C = [X_1, \ldots, X_k]$. We wish to use these values, in an informal manner, to help us to distinguish between two hypotheses. (Note that, for this application, it is more natural to omit the unit constant X_0 from the base C.)

Suppose that we make a belief specification under each hypothesis as follows. We assign a prevision to each member of the collection C and a prevision for each product of two elements of C. (Note that some elements of C may be functionally related to other elements. If we include all functional forms, then this is equivalent to a full probabilistic specification. For example, in the discrete case, each X_i might be the indicator for an elementary event from a partition.)

This determines two inner products over the vector space L of linear combinations of elements of C, namely $(X, Y)_i = P_i(XY)$, where P_i is the prevision assignments under hypothesis i, $i = 1, 2$.

Let T denote the belief transform for $(\cdot, \cdot)_1$ associated with $(\cdot\ \cdot)_2$. Thus T satisfies the relations, for all X, Y in L, that

$$(X, Y)_2 = (X, TY)_1.$$

Denote by $Z_1, \ldots, Z_k$ the orthonormal eigenvectors of T, with ordered eigenvalues $\lambda_k \geq \cdots \geq \lambda_1 \geq 0$. Denote $V_i = Z_i^2$. The eigenvectors and eigenvalues summarise all of the implications of the collection of judgements. For those elements with λ_i large/small, we assign a larger/smaller prevision for V_i under the second hypothesis than under the first hypothesis, while for those elements Z_i with λ_i near 1, we assign a similar prevision for V_i under each hypothesis.

Further this partition is the best achievable, in the sense that Z_1 maximises over L the ratio $P_2(Z^2)/P_1(Z^2)$, Z_2 maximises this ratio over elements of L which are orthogonal to Z_1 and so forth.

Thus an obvious display is to compare the observed Z_i or V_i values with the eigenvalues of T. Under the first hypothesis, all of the V_i values are expected to have a value of 1. Under the second hypothesis, each V_i is expected to have

value equal to λ_i. Thus we may examine the values and informally decide whether either or neither hypothesis seems reasonably supported. Our comparison is analogous to the use of sufficient statistics, in that the information distinguishing the hypotheses is contained in the informative eigenvectors, while "goodness-of-fit" information is carried by the noninformative eigenvectors.

NOTE 1. In principle, the comparison may be over a very large collection of quantities. For example, under a full probability specification, all functional forms of all combinations of sample observations could be compared. However, this would usually be an analytical overkill, and we might select some small subset of functional comparisons, guided by the types of prediction for which we are most concerned that the model should prove satisfactory.

NOTE 2. To simplify the interpretation of the output, we may centre each element X in the base C to have prior prevision 0 under hypothesis 1, yielding a set of normalised eigenvectors of T with zero prevision, unit variance and zero correlation under H_1.

NOTE 3. We have observed that the collection of noninformative eigenvectors provides an informal "goodness-of-fit" test for the models. The eigenvectors are only defined up to an orthonormal transform. However, if there are r orthonormal noninformative eigenvectors, Z_i, then the mean of the square values, $M = (\Sigma Z_i^2)/r$, will have the same numerical value however we choose the eigenvectors. Therefore we may compare the numerical value of M with the prevision for this value under both hypotheses, namely unity.

2.3. *Examples.* We now derive the belief transform for some simple comparisons.

2.3.1. *Changes in location.* Many standard types of hypothesis comparison are concerned with location changes. For the simplest location change, the variance matrix for the elements of C takes a common value under each hypothesis. Centre each X in C so that $P_1(X) = 0$. Each eigenvector Z_i of T satisfies the relation

$$P_2(Z_i^2) = (Z_i, Z_i)_2 = \lambda_i(Z_i, Z_i)_1 = \lambda_i \operatorname{var}_1(Z_i).$$

The variance matrix is common to H_1 and H_2, so that

$$\operatorname{var}_2(Z_i) = \operatorname{var}_1(Z_i) = 1,$$

and we have $P_2^2(Z_i) = (\lambda_i - 1)$, so we can plot the observed value of each Z_i against the appropriate root of $\lambda_i - 1$. In this case, our grid $Z_1, Z_2, \ldots$ has the property that each Z_i is chosen to maximise the absolute value of $P_2(Z)$ over all elements of L which, under the first specification, have zero prevision, unit variance and are uncorrelated with $Z_1, Z_2, \ldots, Z_{i-1}$.

2.3.2. *A simple hypothesis on means.* Suppose that $Y_1, Y_2, \ldots$ form a (in principle infinite) second-order exchangeable sequence. That is, $P(Y_i)$ and $P(Y_i^2)$ do not depend on i, and $P(Y_iY_j)$ for $i \neq j$ does not depend on i or j.

We now describe two possible belief specifications for this sequence. For hypothesis H_k, $k = 1, 2$, we denote the values of $P(Y_i)$, $P(Y_i^2)$ and $P(Y_iY_j)$ for $i \neq j$ by m_k, v_k, c_k, respectively.

Suppose that the current prevision for each Y_i is 0 under each hypothesis, that is, $m_1 = m_2 = 0$. Suppose further that:

Under H_1, the sequence is uncorrelated, that is, $c_1 = 0$.

Under H_2, there is a positive correlation between the members of the sequence, that is, $c_2 > 0$.

Informally, the difference between H_1 and H_2 is that observation of a sample of Y values may change the prevision for future Y_i values under H_2 but not under H_1. The comparison between H_1 and H_2 is essentially the comparison between belief in a zero population mean and belief in a nonzero (but unknown) population mean. The comparison is formally expressed through exchangeability relations. It is often appropriate when we only wish to express beliefs about observable quantities and we only wish to make a small number of belief specifications.

Suppose that we can observe n members, $Y = [Y_1, \ldots, Y_n]$, of the sequence. We construct two inner products over the linear space, L, generated by Y. The inner products describe beliefs under each H_i, that is, $(Y_i, Y_j)_r = \mathrm{Cov}_r(Y_i, Y_j)$, $r = 1, 2$, so that

$$(Y_i, Y_j)_1 = \begin{cases} v_1, & i = j, \\ 0, & \text{otherwise}, \end{cases}$$

$$(Y_i, Y_j)_2 = \begin{cases} v_2, & i = j, \\ c_2, & \text{otherwise}. \end{cases}$$

Thus the matrix representation, U, of the inner product $(\cdot, \cdot)_1$, with respect to basis Y, is

$$U = v_1 I,$$

where I is the identity matrix, and the representation, V, of inner product 2, with respect to Y, is

$$V = w_2 I + c_2 E,$$

where $w_2 = v_2 - c_2$ and E is the matrix all of whose entries are 1. The belief transform for $(\cdot, \cdot)_1$ associated with $(\cdot, \cdot)_2$ has matrix representation

$$T = U^{-1}V = v_1^{-1}(w_2 I + c_2 E).$$

The eigenstructure of T is as follows. The largest eigenvalue of T is

$$\lambda_1 = (w_2 + nc_2)/v_1,$$

corresponding to the normalised eigenvector

$$Z_1 = (n/v_1)^{1/2} Y_M,$$

where $Y_M = (Y_1 + \cdots + Y_n)/n$ is the sample mean.

Thus the single linear combination whose variance differs the most between the two specifications is Z_1, the normalised sample mean, the variance being 1 and λ_1 under hypotheses 1 and 2, respectively. Large values of Z_1^2 therefore "support" hypothesis 2.

The remaining eigenvalues are

$$\lambda_2 = \lambda_3 = \cdots = \lambda_n = (v_2 - c_2)/v_1,$$

corresponding to any $(n - 1)$ mutually uncorrelated standardised combinations $a_1 Y_1 + \cdots + a_n Y_n$ which are also uncorrelated with $(Y_1 + \cdots + Y_n)$, that is, for which $a_1 + \cdots + a_n = 0$. A simple choice is the "cumulative residuals" $R_j = Y_j - (Y_1 + \cdots + Y_{j-1})/(j - 1)$, normalised to variance 1 under hypothesis 1.

(We could alternatively evaluate the usual residuals $Y_j - Y_M$, as these would equivalently span the required space, but for small samples we might need to take account of the correlation between the residuals.)

Thus, if w_2 and v_1 are reasonably different, then we might plot the cumulative or simple normalised residuals to see whether they appear consistent with a variance specification of 1 or w_2/v_1, or neither. The quantitative feature of the plot is whether we have got the variance specifications of about the right order of magnitude. The qualitative feature is whether the residuals can all be reasonably thought to derive from a sequence with similar variance for each value. The qualitative features of the analysis do not depend on the specific numerical inputs, so that, if the hypotheses are determined to be qualitatively of the form that we have discussed, then we may examine the plots without necessarily being very precise in our quantifications.

NOTE 1. In the special case where $w_2 = v_1$, all of the information to distinguish the two hypotheses is in Y_M, and the remaining directions have a strictly diagnostic function.

In our formulation, w_2 represents the residual variation in each observation which cannot be removed by further sampling [see the detailed discussion in Goldstein (1986)]. Thus $w_2 = v_1$ when the residual variation is the same under H_1 and H_2. In particular, this specification applies to the standard location shift problem, in which under H_2 a certain (positive or negative, fixed but unknown) amount has been added to each quantity. This is a special case of the analysis in Section 2.3.1, with $(\lambda_1 - 1)$ nonzero, and all other $(\lambda_i - 1)$ values 0, for which Z_1 is the single diagnostic quantity.

2.3.3. *Testing for a trend.* As an alternative to the above hypotheses, let us suppose that the observations Y_i occur at time points t_i, and that we suspect that there might be a trend in the values across time. Let t denote the column vector $(t_1, \ldots, t_n)^T$.

We introduce hypothesis H_3: Each Y_i can be described by the simple regression model of the form

$$Y_i = m + bt_i + e_i,$$

where m, b are unknown constants, and $e_1, e_2, \ldots$ are a second-order exchangeable sequence of random quantities, each with prevision 0.

We rewrite hypotheses H_1 and H_2 of the preceding section in the notation of this model by the following assignments:

Under H_1: $Y_i = e_i$, that is, $P(m) = \text{var}(m) = P(b) = \text{var}(b) = 0$.
Under H_2: $Y_i = m + e_i$, that is, $P(m) = P(b) = \text{var}(b) = 0$.

Suppose that under H_1, H_2, H_3 the same value, v_e, is assigned for the variance of each e_i. (This implies, in the notation of the preceding section, that $w_2 = v_1$, i.e., that all but one of the eigenvalues of the transform between H_1 and H_2 have value 1.)

Denote by u_m, v_m the variance of m under H_2 and H_3, respectively. Denote by v_b the variance of b under H_3. Suppose, for simplicity, that the origin for time has been chosen so that $(t_1 + \cdots + t_n) = 0$, and that the covariance between m and b under H_3 is 0.

The matrix representations V_i for the three inner products $(X, Y)_i = P_i(X, Y)$, $i = 1, 2, 3$, are as follows:

$$V_1 = v_e I,$$
$$V_2 = v_e I + u_m E,$$
$$V_3 = v_e I + v_m E + w_b G,$$

where $G = tt^T$ and $w_b = v_b + P_3^2(b)$.

There are three comparisons that we might make:

H_1 against H_2, a simple location shift alternative;
H_2 against H_3, a trend test given nonzero, but unknown, level;
H_1 against H_3, comparing a zero-mean, exchangeable sequence to trend plus general level.

Let T_{ij} be the transform for H_i associated with H_j.

We compared H_1 to H_2 in the preceding section. We now compare H_2 with H_3. The matrix form for T_{23} is

$$T_{23} = V_2^{-1} V_3 = (1/v_e)(I - kE)(v_e I + v_m E + w_b G),$$

where $k = u_m/(nu_m + v_e)$, so that, as $EG = 0$,

$$T_{23} = I + dE + rG,$$

where $d = ((v_m - u_m)/(nu_m + v_e))$ and $r = (w_b/v_e)$.

There are two informative eigenvectors of T_{23}. The first is W_M $[= (t_1 Y_1 + \cdots + t_n Y_n)/n]$, the sample covariance of t and Y, with eigenvalue $\lambda_1 = 1 + t_{(2)} r$, where $t_{(2)} = (t_1^2 + \cdots + t_n^2)$. The second is Y_M, the sample mean, with eigenvalue $\lambda_2 = 1 + nd$.

NOTE 1. If the uncertainty for the level is the same under H_2 and H_3, that is $v_m = u_m$, then $\lambda_2 = 1$, and the only informative eigenvector is W_M.

NOTE 2. As before, the least squares residuals $R_i = Y_i - Y_M - b^* t_i$, where $b^* = W_M / t_{(2)}$, span the space of unit eigenvectors, and may thus be plotted for diagnostic purposes, though if n is small we might again choose to work with the corrected residuals.

NOTE 3. Again, the qualitative features of the analysis do not depend on the precise numerical values, so that as before the influence of the belief specification is firstly to direct us qualitatively to the kind of features of the data that we should examine and secondly to give us quantitative guidelines to compare with our qualitative assessments.

NOTE 4. Given a nonzero covariance between b and m under H_3, there would again be two informative eigenvectors of T. In this case, each would be a linear combination of vectors W_M and Y_M.

Finally, we compare H_1 with H_3. We can form this comparison directly from the comparisons H_1 with H_2 and H_2 with H_3, by the relation

$$T_{13} = T_{12} T_{23}.$$

As the space, M, spanned by Y_M and W_M is sufficient for transforms T_{12} and T_{23}, M is also sufficient for T_{13}. (This is a special case of the corollary of Section 1.3.)

We can form a basis, consisting of Y_M, W_M, and any $(n - 2)$ combinations orthogonal to both Y_M and W_M. This basis is an eigenbasis for each of T_{12}, T_{23} and T_{13}. All the eigenvectors of T_{13} are noninformative except Y_M for which the eigenvalue is $(1 + n(v_m/v_e))(1 + nd)$, and W_M with eigenvalue $1 + t_{(2)}r$. Thus the values Y_M and W_M may be used to distinguish H_1 and H_3.

However, as we have seen above, W_M distinguishes H_2 from H_3, while, particularly in cases with $d = 0$, Y_M distinguishes H_1 from H_2. Thus splitting T_{13} by introducing H_2 may be useful to clarify which aspects of H_1 are "contradicted" by the data, and, in particular, whether it is sufficient to move from H_1 to H_2 or whether we need to introduce the full H_3.

3. Example.

3.1. *Problem description.* To illustrate the methodology of hypothesis comparison, we give the following example. The situation that we are analysing relates to certain aspects of an industrial smelting process for aluminium. Alumina Al_2O_3 is reduced (i.e., the metal is extracted from the oxide) by electrolysis. A (more or less) constant electric current is passed through a hot solution containing alumina. The current is carried on charged Al^{3+} ions. In principle, we can calculate how much the aluminium is being reduced by Faraday's law. However, there are certain practical complications, namely that

some Al^{3+} recombines with oxygen, and also "current efficiency" is less than 100% (in our case, usually about 80%). As long as current efficiency is independent of alumina concentration, aluminium is produced at a constant rate and hence Al_2O_3 is used at a constant rate, within normal operating conditions. A further complication is that various extraneous features of the process ("sludge," "crust," etc.) all contain undissolved alumina which gradually dissolves or comes out of solution, depending on a variety of unpredictable circumstances. However, experience strongly suggests that the overall trend is still more or less linear.

In a certain experiment, under "normal" operating conditions, the percent-concentration of alumina in the solution was determined every 10 minutes. The experiment was run several times, each time terminating when an "anode effect" occurred (corresponding to the concentration falling to a critical value). As the value of the observation is essentially fixed at the end of the series, it will simplify our description of the model to call the last time point t_1, the second to last time t_2 and so on. We call y_{tr} the alumina concentration at the tth time point from the end, on run r. The interest in this example is to consider the underlying average performance and to study the stochastic behaviour of the system. The physical specification of the problem suggests the following model formulation.

3.2. *Model: simple errors.* The first model used to represent the values of y for run r was

$$y_{tr} = \text{intercept}_r + \text{slope}_r t + \text{error}_{tr}, \tag{3.2.1}$$

where slope_r and intercept_r are the two constants for run r which define the underlying deterministic physical process of extraction of the metal and are considered exchangeable between runs. The error term expresses the various discrepancies from the linear trend. For our first analysis, we will suppose that this error represents the pure measurement error for the reading y_{tr}. This is not a particularly realistic supposition, and one of the purposes of the study is to examine the stochastic behaviour of the error terms. However, this simple specification will allow us to apply the analysis of the preceding section directly to the deterministic components of the equation. (We will analyse the error term in more detail in Section 3.6.) Thus, in our initial analysis, all error quantities are considered exchangeable, uncorrelated with each other and all other quantities and to have prior prevision 0.

3.3. *Belief transform.* We compare the two hypotheses H_2 and H_3 which were described in Section 2.3.3. In the notation of that section, we evaluate the operator T_{23}, for the trend test given unknown level, that is, for comparing H_2 against H_3, over the linear space spanned by $Y_1, \ldots, Y_n$.

For simplicity, we suppose that the prevision and variance for the "intercept" term is the same under the two hypotheses. We could reform the model so that the mean time was 0, but it is more natural in this example to have time points $t = 1, 2, 3, \ldots, n$. In the notation of Section 2.3, the operator

becomes

$$T_{23} = I + rG - sH,$$

where $H = ut^T$, $u = (1, 1, \ldots, 1)^T$, and $s = ku^T t w_b / v_e$. Thus T_{23} has a single informative eigenvector, proportional to

$$Z_1 = r(t_1 Y_1 + \cdots + t_n Y_n) - s(Y_1 + \cdots + Y_n) = d_1 Y_1 + \cdots + d_n Y_n,$$

where $d_i = rt_i - s$, with eigenvalue

$$\lambda = 1 + r(t_1^2 + \cdots + t_n^2) - s(t_1 + \cdots + t_n).$$

We correct all of the data, by subtracting the prior mean of each Z_i under H_2, or equivalently by subtracting from each Y_i the prior mean of Y_i under H_2. This prior mean is equal to P(intercept) for each i. Under H_2, we have laid an uncorrelated grid, of 13 directions, on the Y space, where each Z_i has zero prevision and unit variance. Under H_3, Z_1 has $P(Z_1^2) = \lambda$, and each other $P(Z_i^2) = 1$. Thus the value of Z_1^2 is informative for distinguishing between the hypotheses, while the remaining Z_i^2 values may offer simple indications of degree of fit of the pair of models.

3.4. *Example analysis: data and beliefs.* We will analyse a set of three runs, each under "normal" operating conditions. Two of the series (the first and third) ran for 18 observations each, while the other ran for 13 observations. To simplify our account, we evaluate the same transform for each series. Therefore we truncate the longer series to 13 values, so that each series corresponds to the last 13 values. The data that we use in this analysis are listed in Table 1.

TABLE 1

Percent-concentration of alumina in solution, determined every 10 minutes for three runs, each under "normal" operating conditions (the observations are presented in reverse time order)

Run 1	Run 2	Run 3
1.79	1.93	1.54
2.14	1.76	1.48
2.13	1.61	1.57
2.07	2.32	1.28
2.08	1.87	1.50
1.88	1.80	1.79
1.94	2.21	1.88
2.01	2.23	2.11
2.35	2.42	2.48
2.23	2.58	2.28
2.58	2.60	3.39
2.48	2.65	3.44
2.82	2.70	2.80

We will analyse the three runs under the following belief specifications:

		H_2	H_3
$P(\text{intercept}_r)$	=	1.4	1.4
$P(\text{slope}_r)$	=	0	0.1
$\text{var}(\text{intercept}_r)$	=	0.058	0.058
$\text{var}(\text{slope}_r)$	=	0	0.0017
$\text{cov}(\text{intercept}_r, \text{slope}_r)$	=	0	0
$\text{var}(\text{error}_{tr})$	=	0.04	0.04

(From the exchangeability specification, these assessments are the same for all t and r.)

While these specifications are intended simply as an illustration of the types of analysis that we might perform, the above values did actually represent the belief specifications of the analyst and were based upon both theoretical arguments and pilot studies under similar operating conditions. The sole exception is the error variance, which has been set at the actual assessed marginal value, but for which the errors were not actually judged exchangeable—we return to this in Section 3.6.

(I am very grateful to Malcolm Farrow for explaining to me the above problem and for providing the data and belief specifications that we are using. Notice that the specified variances are small, corresponding to a high prior confidence that the stochastic development is roughly as specified, with underlying values for which we have fairly reliable prior estimates. As a general observation, precise prior inputs provide clear distinctions between the competing models, while being very sensitive to "small" discrepancies from our descriptions.)

The analysis of this collection of data involves a further level of belief specification, namely an exchangeability specification for the slope and intercept coefficients across runs, in the manner of Goldstein (1986). However, for the purpose of this illustration, rather than introducing another layer of complexity into the analysis, we follow the alternative of analysing each series separately. We then compare the analyses to identify informally the differences between the series. Formal analyses relating the individual series involve belief transforms for exchangeable systems, which we will consider elsewhere.

Thus we shall evaluate the transform for comparing the hypotheses based on a single run and use the three runs as repetitions against which we can test our formulation.

3.5. *Example*: *eigenanalysis*. With the above belief specifications, the single informative eigenvalue for T_{23}, as evaluated in Section 3.3, takes value 63.6, with corresponding eigenvector $Z_1 = d_1Y_1 + \cdots + d_{13}Y_{13}$, where d_i is proportional to $0.0428i - 0.285$.

The variance of Z_1 under H_2 is 1, with zero prevision, while $P(Z_1^2)$ under H_3 is 63.6.

TABLE 2

Run	Mean Z_i^2	Standard deviation Z_i^2	Range
1	0.996	0.789	0.010–2.451
2	0.910	1.133	0.002–3.502
3	3.049	6.408	0.000–23.96

(As in section 3.3, we have subtracted from each Y_i the prior prevision under H_2, which in this case has value 1.4.)

For the above data, the observed values of Z_1^2 are as follows:

$$\text{Run 1}, \quad Z_1^2 = 24.4,$$

$$\text{Run 2}, \quad Z_1^2 = 42.3,$$

$$\text{Run 3}, \quad Z_1^2 = 128.8.$$

The average over the three runs agrees almost exactly with the prior expectation under H_3, supporting the alternative of positive slope.

The remaining 12 eigenvectors each have $P(Z_i^2) = 1$, under both H_2 and H_3. We selected a particular grid and evaluated the collection of Z_i^2 values, $i = 2, \ldots, 13$, for each series.

As noted in Section 2.2, $M = \Sigma_2^{13} Z_i^2/12$, is invariant over choices of grid, and values near 1 suggest a reasonable fit. In Table 2, we tabulate the value of M for each run and also give the standard deviation and range of the Z_i^2 for each run over our chosen grid. (For example, analysing the first series, we evaluate 12 Z_i^2 values, the smallest being 0.010, the largest being 2.451 and the average of the 12 values being 0.996.)

The values for the third run were inflated by a single extremely large value, namely 23.96. If we eliminate this value, then the mean of the remaining values falls to 1.148, with standard deviation 1.197 and largest value 4.012. The sample correlations between the Z_i^2 values across the three runs all have absolute magnitude less than 0.2, suggesting no unsuspected consistencies across runs. Informally, our analysis suggests that the simple slope and intercept model makes reasonable predictions across the grid over the series of three runs.

Notice that the single "aberrant" value, namely 23.96, occurred in the same run which had by far the largest observed Z_1^2 value, suggesting possible quantitative differences between run 3 and the first two runs.

The above analysis suggests first that beliefs about the slope are of roughly the right order of magnitude, and secondly that the simple slope/intercept model describes the sequence variability adequately to a "first-order" approximation.

We could introduce formal accept/reject type of decisions based upon our analysis, provided that we were prepared to make higher-order belief specifications in order to judge the "variability" of our eigenvariances. However, this

would take us beyond our current investigation, namely to examine the implications of a fairly minimal collection of belief specifications for distinguishing the hypotheses. Often, as in the above analysis, an informal judgement will be adequate (supported, if necessary by simple probability arguments such as Markov's inequality). In other cases, the conclusion of the analysis must be that the combination of the observed data and the limited belief specification is not sufficient to distinguish the hypotheses, though the analysis will still offer some limited guidance. (For example, both in the above analysis and the analysis of Section 3.7, the third series is repeatedly identified as behaving differently than the first two series.)

We now carry out a more detailed analysis of the error terms in the model.

3.6. *Model: general errors.* A more realistic description of our uncertainties concerning the stochastic development of the process represented the series as follows:

$$y_{tr} = \text{intercept}_r + \text{slope}_r t + e_{tr} + v_{tr} + h_{tr}. \tag{3.6.1}$$

In this equation, e, v and h represent the three types of "error" which are superimposed on the observation.

(i) v_{tr} expresses the stochastic development of the series as a random walk with drift (equal to the slope) so that the successive differences in the series are uncorrelated jumps. We express v_{tr} as

$$v_{tr} = \sum_{i=1}^{t} f_{ir},$$

where all quantities f_{tr} are considered exchangeable, uncorrelated with each other and all other quantities and to have prior prevision 0.

(ii) h_{tr} represents the measurement, in the chemical analysis of the sample, of suspended particles, as well as the dissolved alumina. The amount in suspension will vary, but will have similar effects in neighbouring time points and is expressed by an autoregressive term as

$$h_{tr} = ah_{(t-1)r} + u_{tr},$$

where a is an autoregressive parameter, and all quantities u_{tr} are considered exchangeable, uncorrelated with each other and all other quantities and to have prior prevision 0. There is no u_{1r} term, but h_{1r} has variance chosen so that variances for h_{tr} are constant over t.

(The autoregressive parameter is an aspect of our uncertainty, which must be specified, rather than a physical parameter which must be estimated. Partly, this is for simplicity, but mainly it is because this did actually represent the beliefs of the analyst.)

(iii) e_{tr} represents the pure measurement error for the reading on y_{tr}. All quantities e_{tr} are considered exchangeable, uncorrelated with each other and all other quantities, with prior prevision 0.

The minimal belief specification that we require to determine second-order beliefs for the data requires us to specify the following additional quantities: $\text{var}(e_{tr})$, $\text{var}(f_{ir})$ and $\text{var}(h_{tr})$. (From the exchangeability specifications, these assessments are the same for all t and r.) We must also state a value for the autocorrelation parameter, a.

For our example analysis, we will use the following specifications:

$$\text{var}(e_{tr}) = 0.01, \qquad \text{var}(f_{ir}) = 0.01, \qquad \text{var}(h_{tr}) = 0.0204, \qquad a = 0.7.$$

Note that the marginal error variance for Y_1 is the same under this specification and the specification of Section 3.4, and first differences for the series have constant marginal variance. (These values were thought reasonable from previous studies and theoretical arguments—again, I am grateful to Malcolm Farrow, for providing the model and belief inputs.)

3.7. *Eigenanalysis.* We now compare the two specifications, H, corresponding to model (3.2.1), and H^*, corresponding to (3.6.1), by generating the two corresponding inner products over the base $[Y_1, \ldots, Y_{13}]$, and extracting the 13 eigenvalues and vectors of the corresponding belief transform.

As each Y_i has the same prevision under both H and H^*, we subtract from each Y_j the common prior prevision [namely $P(Y_r) = 1.4 + 0.1r$]. This is equivalent to comparing the inner products over the two spaces generated by the covariances rather than the product moments.

For each individual run of the sequence, the value of Z_i^2 was evaluated for each $i = 1, \ldots, 13$. For this decomposition, each Z_i has prevision 0 and the collection $Z_1, \ldots, Z_{13}$ is uncorrelated under each specification. The prior variance for each Z_i is 1 under H. Each Z_i has maximum variance under H^*, given unit variance under H, and zero correlation with previous Z_i values.

In Table 3, we list the ordered eigenvalues of the transform and the corresponding average of the three evaluations of the value of Z^2. (For example, the largest eigenvalue is 3.23 and the average of the values of Z_1^2 over the three runs is 6.05.)

Note firstly that as none of the eigenvalues are particularly large or small, there is no particular linear combination that will be strongly diagnostic between hypotheses. However, the average values for the two strongest diagnostic directions, Z_1^2 and Z_2^2, appear to favour H^*. Furthermore, the four smallest eigenvalues all correspond to small Z^2 values, but most of the observed eigenstructure could be consistent with either specification. There does not seem to be dramatic evidence from the display which would cause us

TABLE 3

	1	2	3	4	5	6	7	8	9	10	11	12	13
Eigenvalue	3.23	2.37	1.55	1.42	1.17	0.93	0.77	0.66	0.60	0.55	0.52	0.50	0.49
Average Z^2	6.05	3.06	0.90	0.58	0.22	1.12	1.82	1.80	2.40	0.57	0.87	0.70	0.70

to reject the pair of models. We might feel, from the values for the largest eigenvalues, that there is somewhat more support for H^* than for H.

The largest Z^2 value, namely 6.05, may be worth describing further. The corresponding eigenvector, describing the sharpest "nonlinear" error effect that we expect, a priori, under the general error model, was

$$Z_1 = -0.86Y_1 - 0.61Y_2 - 0.18Y_3 + 0.32Y_4 + 0.77Y_5 + 1.1Y_6 + 1.1Y_7 + 0.89Y_8 + 0.49Y_9 - 0.02Y_{10} - 0.51Y_{11} - 0.85Y_{12} - 0.97Y_{13}.$$

The three values of Z_1^2 were 2.80, 0.54 and 14.75, for series 1, 2 and 3, respectively. (The value of 14.75 is the only Z_i^2 value which is larger than 5 for any of the three series.) As in the analysis of Section 3.5, we identify the third series as being consistently highly variable with respect to our specifications.

4. Concluding comments.

4.1. *On hypothesis comparisons.* We have suggested a general methodology for comparing two, or several, hypotheses. This methodology is of particular value when we do not wish to make a full specification of beliefs. For example, in Section 3.5, we demonstrated strong support for belief in a linear trend term in the data, with very modest belief specifications. Again, for the model of Section 3.6, any reasonable, full probability specification would prove difficult both to specify and to analyse.

In our approach, particular belief statements about model quantities imply various belief statements about certain observable quantities which are compared with the actual behaviour of the observed quantities. We could make more detailed belief statements about the quantities, if we were both willing and able to do so, and then build our hypothesis comparison over the larger collection of beliefs. The difficulty with exclusive reliance on likelihood-ratio-type comparisons is that there is no way to give "easy answers to easy questions." We are forced into extremely detailed levels of belief specification, not as a matter of choice but as a matter of necessity. Often this results in a high degree of arbitrariness in the specification. In such cases, it is very hard to interpret a likelihood-ratio-type comparison, as there is no way to distinguish those aspects of the comparison which relate to meaningful belief specifications from those which are purely artifacts of the analysis.

Because the method of hypothesis comparison does not lead unambiguously to an accept/reject type of decision, the approach may be considered complementary to such formal analyses. If the data do support one of the hypotheses strongly, then we would expect that a comparison of the observed behaviour of the system, of the type that we have suggested, over natural subspaces of interest, should lead to similar conclusions to those of the full analysis. Indeed, if the favoured model does not make better predictions for the data, then this may call into question the apparent conclusions of the formal analysis.

We have emphasised the informal nature of the hypothesis comparison. However, because the hypothesis comparison separates out the differences

between the predictions under the differing hypotheses in a very efficient manner, the transform also plays an important role in formal inferential procedures based on the belief specifications, essentially allowing us to separate beliefs over the hypotheses from the beliefs given the hypotheses. We will discuss this in detail elsewhere.

4.2. *On belief transforms.* We have described the method of hypothesis comparison in some detail, partly because it is useful in its own right, but mainly to illustrate the basic argument of this paper, namely that it is very often useful to be able to systematically compare various collections of belief specifications. This comparison can be made automatically through the associated belief transform which exists in full generality. The hypothesis comparison is the simplest type of belief specification that we can make, as various initial belief specifications are directly compared. In general we will compare derived belief measures, such as expected information for various competing experimental designs or sampling frames, typically involving perturbations of an initial belief specification. The derivation of the derived belief structures may be technically complicated, but in all cases the comparison of the beliefs will follow automatically from the associated belief transform, which unifies all of the apparently different forms of analysis. We will discuss the treatment of such general types of belief transform elsewhere.

Acknowledgments. All of the computations in Section 3 were made with the package [B/D] (an acronym for beliefs adjusted by data), which provides a computer language for analysing belief structures. This program has been developed by David Wooff and myself, under a grant from the Science and Engineering Research Council.

REFERENCES

BACHMAN AND NARICI (1966). *Functional Analysis*. Academic, New York.

DE FINETTI, B. (1974). *Theory of Probability* **1**. Wiley, New York.

GOLDSTEIN, M. (1981). Revising previsions: A geometric interpretation. *J. Roy. Statist. Soc. Ser. B* **43** 105–130.

GOLDSTEIN, M. (1986). Exchangeable belief structures. *J. Amer. Statist. Assoc.* **81** 971–976.

GOLDSTEIN, M. (1987a). Systematic analysis of limited belief specifications. *The Statistician* **36** 191–199.

GOLDSTEIN, M. (1987b). Can we build a subjectivist statistical package? In *Probability and Bayesian Statistics* (R. Viertl, ed.) 203–217. Plenum, New York.

GOLDSTEIN, M. (1987c). [B/D]: Introduction and overview. Technical report, Sch. Mathematics, Univ. Hull.

GOLDSTEIN, M. (1988a). Adjusting belief structures. *J. Roy. Statist. Soc. Ser. B* **50** 133–154.

GOLDSTEIN, M. (1988b). The data trajectory. In *Bayesian Statistics 3* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 189–209. Oxford Univ. Press.

WOOFF, D. A. (1987). [B/D]: Reference manual. Technical report, Sch. Mathematics, Univ. Hull.

DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF DURHAM
DURHAM, DH1 3LE
ENGLAND