The Annals of Statistics
1993, Vol. 21, No. 3, 1119–1138

FIDUCIAL PREDICTION AND SEMI-BAYESIAN INFERENCE

BY A. PHILIP DAWID AND JINGLONG WANG

University College London and East China Normal University

We investigate the problem of fiducial prediction for unobserved quantities within the framework of the *functional model* described previously by Dawid and Stone. It is supposed that these are related to a completely unknown parameter by means of a regular functional model, and that the observations are either given as known functions of the predictands, or are themselves related to them by means of a functional model. We develop algebraic conditions which allow the application of fiducial logic to the prediction problem, and explore the consequences of such an application—some of which appear unacceptable unless still stronger conditions are imposed.

A reinterpretation of the fiducial prediction problem is given which can be applied to yield an inferential distribution for the unknown parameter in the presence of partial prior information, expressible as a functional hypermodel for the parameter, governed by a completely unknown hyperparameter. This solution agrees with the fiducial distribution when the hypermodel is vacuous and with the Bayes posterior distribution when the hyperparameter is fully known, but allows in addition for intermediate levels of partial prior knowledge.

1. Introduction. Dawid and Stone (1982) introduced the general concept of a *functional model*, expressing an *observable* X as a function of an unknown *parameter* Θ, and an *error* variable E having a known distribution, independent of the value of Θ. This notion encompasses the *structural model* of Fraser (1968), and is closely related to the *pivotal model* of Barnard (1980, 1985). Dawid and Stone (1982) showed how, under some additional algebraic conditions, functional models (called in this case *regular*) can support a form of *fiducial inference* about Θ. However, no claim was made that this framework is either necessary or sufficient for the drawing of valid fiducial inferences. Indeed, it was shown that, without further conditions, an inconsistency can arise between two possible methods, both apparently reasonable, of conditioning in a fiducial distribution. The purpose of the analysis of Dawid and Stone (1982) was not to recommend fiducial inference, but to explore its properties, both good and bad, within the functional model framework.

In this paper we extend that exploratory analysis to cover *fiducial prediction* of unobserved random quantities. Special cases of fiducial prediction have previously been considered by Fisher [(1935), (1973), pages 117–122], Kitagawa (1957) and Hora and Buehler (1967). Here we identify and analyse the additional algebraic structure needed in order that predictive fiducial inferences may be drawn within the functional framework, and we explore some of the properties of the resulting predictive distributions. Again, some of

Received September 1989; revised October 1992.
AMS 1991 *subject classifications*. Primary 62A30; secondary 62A15.
Key words and phrases. Fiducial inference, functional model, prediction, Bayesian inference.

these are counterintuitive, suggesting the need for further restrictions on the application of fiducial logic.

Section 2 gives a brief résumé of the underlying structure, as described by Dawid and Stone (1982), with which we work, and Sections 3 and 4 develop some necessary algebraic apparatus. In particular, in Section 3 we investigate necessary and sufficient conditions for a function Y of X to be describable by a regular functional model when X is; while in Section 4 we demonstrate how any regular functional model can be decomposed into a "basic" model and an ancillary part.

In Section 5 we study fiducial prediction when the observed variable Y is a function of the full predictand X. We also construct a joint fiducial distribution for (X, Θ) given Y, and note that further conditioning of this distribution on Θ can give an answer at variance with the sampling distributions for X given Y and Θ, unless the functional model for Y has the special property of being *pivotal*. In Section 6 we turn to the more general case of a stochastic relationship of Y to X, itself described by means of a functional model. Then Section 7 takes up the study of fiducial prediction for X based on Y in such a model. Here there are seemingly two possible approaches, the "global" and the "stepwise," which generally disagree; however, we argue that there is no inconsistency here, since only the global approach takes into account all the available information. Finally, Section 8 applies these results to develop a methodology of "semi-Bayesian" inference when there is a hierarchical prior structure. The Appendix collects together some purely algebraic proofs.

The main findings of our analysis are, in summary, that whilst the logic of fiducial prediction within regular functional models possesses a good measure of internal consistency, it can deliver answers at odds with sampling behaviour. Similar warnings, from a different standpoint, have been sounded by Stone (1976).

2. Background.

We recall the following background from Dawid and Stone (1982), which should be consulted for further details. A *functional model* relates three quantities: an *observable*, X, taking values in a measurable space $(\mathscr{X}, \Sigma_{\mathscr{X}})$; a *parameter* Θ, with values in $(\boldsymbol{\Theta}, \Sigma_{\boldsymbol{\Theta}})$; and an *error*, E, with values in $(\mathscr{E}, \Sigma_{\mathscr{E}})$. We shall largely ignore the (nontrivial) technicalities associated with measurability questions, and simply assume that all sets and functions introduced are suitably measurable. We suppose that E is a random variable with a specified distribution P over $\mathscr{E}$, independent of the value of Θ. The relationship between these quantities is specified by a known surjective function $\boldsymbol{\Theta} \times \mathscr{E} \to \mathscr{X}$; we do not introduce a special name or notation for this function, but denote its value at (θ, e) simply by θe (or $\theta \circ e$, $\theta * e$ etc. when we need to distinguish different functions). The functional model is then defined by the relationship $X = \Theta E$; it may be denoted by $\langle X = \Theta E,\ E \sim P \rangle$ or just $\langle X = \Theta E \rangle$.

For any $\theta \in \boldsymbol{\Theta}$, we can identify θ with the function $\mathscr{E} \to \mathscr{X}$ given by $e \mapsto \theta e$. It is clear that any value θ for Θ induces a probability distribution P_θ

for X, viz. that of θE when $E \sim P$, and hence that the functional model induces a statistical model $\mathscr{P} = \{P_\theta: \theta \in \mathbf{\Theta}\}$.

This interpretation of $\mathbf{\Theta}$ as a set of transformations on $\mathscr{E}$ is fairly standard, for example, in Fraser's *structural model* [Fraser (1968)], where the set is further required to form a transformation group. However, for our purposes it will prove more helpful to identify each $e \in \mathscr{E}$ with the function $\mathbf{\Theta} \to \mathscr{X}$ given by $\theta \mapsto \theta e$, thus regarding $\mathscr{E}$ as a set of transformations (written on the right) from $\mathbf{\Theta}$ to $\mathscr{X}$.

We call $x \in X$, $e \in \mathscr{E}$ *compatible* if $x = \theta e$ for some $\theta \in \mathbf{\Theta}$, and write $\mathscr{E}_x$ for the set of $e \in \mathscr{E}$ compatible with $x \in X$. The functional model is *invertible* if, for any compatible pair (x, e), the solution θ of $x = \theta e$ is unique. We then express this solution as $\theta = xe^{-1}$. The model is *simple* if such a solution exists for all $x \in X$, $e \in \mathscr{E}$; or, more generally, *partitionable* if, for any $x_1, x_2 \in \mathscr{X}$, $\mathscr{E}_{x_1}$ and $\mathscr{E}_{x_2}$ are either identical or disjoint. Partitionability is equivalent to the existence of surjective functions $a: \mathscr{X} \to \mathscr{A}$ and $u: \mathscr{E} \to \mathscr{A}$, such that x and e are compatible if and only if $a(x) = u(e)$. The functions a and u are determined up to an invertible transformation. We call $A = a(X) = u(E)$ the *functional ancillary*. Defining the *canonical reduction* a^* by $a^*(x) = \mathscr{E}_x$, partitionability thus holds if the range $\mathscr{A}^*$ of a^* is a partition of $\mathscr{E}$; and one possible choice for the functional ancillary is then $A^* = a^*(X)$, which we then term the *canonical ancillary*.

A functional model which is both invertible and partitionable will be termed *regular*; if, moreover, it is simple, it will be called *basic*. Partitionability is a fundamental algebraic requirement on a model if the observation of a value x for X is to be used to update the distribution of E, for then and only then has the logical information obtained about E the form of an observed value $[a(x)]$ for a function of E $[u(E)]$. One can then argue that one should use, as the appropriate inferential distribution of E after observing $X = x$, its conditional distribution given the observed information $u(E) = a(x)$. Invertibility ensures that, given the data x, Θ is a function of E. This function then induces, from the inferential distribution of E, an inferential distribution—the *fiducial* distribution—for Θ. Thus, in the regular case, the fiducial distribution of Θ is that induced by the *fiducial model* $\Theta = xE^{-1}$, where now the distribution of E is that calculated from its unconditional distribution P by conditioning on the observed value $a(x)$ of $u(E)$. [For a basic model $u(E)$ is trivial, and no conditioning is needed.]

The logic of the above "fiducial inversion," which is similar to that propounded by Fisher, has (to say the least) not met with universal acceptance. The principal assumption to which objection can be made is that it is appropriate to use, as an inferential distribution for E based on data $X = x$, its conditional distribution given $u(E) = a(x)$. While it is true that this information exhausts all the *logical* knowledge of E conveyed by observing $X = x$, it is not clear that no additional (perhaps probabilistic) information is conveyed. For this reason, Fisher, Fraser and other proponents of such fiducial arguments have been at pains to limit the applicability of fiducial logic to cases in which there is, in some intuitive sense, "no prior information" about Θ. Our attitude towards its applicability even in that case remains noncommittal, beyond considering that the argument has enough prima facie appeal to make it worthwhile to investigate its implications. These, in turn, may shed more light on its acceptability. This is the programme initiated by Dawid and Stone (1982), and continued here.

3. Regular contractions. In this section we consider algebraic conditions under which the regularity of a functional model is preserved on replacing the observable X by a function Y of X (so that fiducial inference for Θ is still available if only Y is observed.) All proofs are confined to the Appendix.

Suppose then that we have a regular functional model (RFM) M_1: $\langle X = \Theta E\rangle$, with functional ancillary $A = a(X) = u(E)$. Let k be a surjective function from $\mathscr{X}$ to some space $\mathscr{Y}$, and define a new operation $*$ by $\theta * e = k(\theta e)$. Let $Y = k(X)$. Then $Y = \Theta * E$. The new model M_2: $\langle Y = \Theta * E,\ E \sim P\rangle$ is the *contraction* of M_1 defined by k. In general, M_2 need be neither invertible nor partitionable. If it is both, so that M_2 is regular, we call it a *regular contraction* (or a *basic contraction* if M_2 is basic).

Examples. For Examples 3.1 and 3.2, we take M_1 to be the RFM with $\mathscr{X} = \mathscr{E} = \mathbb{R}^n$, $n > 1$, $\boldsymbol{\Theta} = \mathbb{R}^1$, $X = (X_1, \dots, X_n)$, $E = (E_1, \dots, E_n)$, and location structure $X_i = \Theta + E_i$, $i = 1, \dots, n$. A possible choice for A is $(X_2 - X_1, \dots, X_n - X_1) = (E_2 - E_1, \dots, E_n - E_1)$.

Example 3.1. Let $\mathscr{Y} = [0, \infty)$, $Y = \sum_{i=1}^n X_i^2$, yielding reduction M_2: $\langle Y = \Theta * E\rangle$, where $\theta * e = \sum_{i=1}^n (\theta + e_i)^2 = n(\theta + \bar{e})^2 + s_e^2$, with $\bar{e} = \sum_{i=1}^n e_i/n$, $s_e^2 = \sum_{i=1}^n (e_i - \bar{e})^2$. Given $y \in \mathscr{Y}$, $e \in \mathscr{E}$, there exists no value $\theta \in \boldsymbol{\Theta}$ with $y = \theta * e$ if $y < s_e^2$, just one such value if $y = s_e^2$, and two distinct values if $y > s_e^2$. In particular, M_2 is not invertible. Since the condition that y and e be compatible is that $y \geq s_e^2$, and as y varies the sets $\{e: y \geq s_e^2\}$ do *not* constitute a partition, M_2 is not partitionable.

Example 3.2. Let $1 \leq m < n$ and take $Y = (Y_1, \dots, Y_m)$ with $Y_i = X_i/s_X$. Thus $\theta * e = ((\theta + e_i)/s_e: i = 1, \dots, m)$. (We should really remove from $\mathscr{E}$ the set $\{e: e_1 = e_2 = \cdots = e_n\}$, and similarly for $\mathscr{X}$. With a continuous distribution for E, this is of no consequence.) Then, for $y \in \mathscr{Y}$ and $e \in \mathscr{E}$, $y = \theta * e \Rightarrow y_i - y_1 = (e_i - e_1)/s_e$, $i = 2, \dots, m$. Conversely, if this condition holds, $y_i = y_1 + (e_i - e_1)/s_e = (\theta_0 + e_i)/s_e$, $i = 1, \dots, m$, where $\theta_0 = y_1 s_e - e_1$, so that $y = \theta_0 * e$. The condition for y and e to be compatible is thus that $y_i - y_1 = (e_i - e_1)/s_e$, $i = 2, \dots, m$, which shows that M_2 is partitionable. (If $m = 1$, M_2 is simple.) Furthermore, $\theta_1 * e = \theta_2 * e \Rightarrow (\theta_1 + e_1)/s_e = (\theta_2 + e_1)/s_e \Rightarrow \theta_1 = \theta_2$, so that M_2 is invertible. Hence M_2 is regular (basic if $m = 1$). Note that M_2 is not a "location model."

We shall now investigate when a contraction M_2 of M_1, defined by a function k: $\mathscr{X} \to \mathscr{Y}$, will be invertible, partitionable or simple.

THEOREM 3.1. *M_2 is invertible if and only if*

$$\{a(x_1) = a(x_2) \text{ and } k(x_1) = k(x_2)\} \Rightarrow x_1 = x_2.$$

[*In this case we call the functions $a(\cdot)$ and $k(\cdot)$* transverse.]

PROOF. See the Appendix.

In order to investigate the partitionability of M_2, we need some general notation. For surjective functions $f\colon \mathscr{S} \to \mathscr{F}$ and $g\colon \mathscr{S} \to \mathscr{G}$, write $f \le g$ if there exists a function $\phi\colon \mathscr{G} \to \mathscr{F}$ such that $f = \phi \circ g$ [i.e., $f(s) = \phi(g(s))$, all $s \in \mathscr{S}$]. In this case f is a *reduction* of g. The function ϕ is uniquely determined. We write $\phi = f \circ g^{-1}$.

Given any $f\colon \mathscr{S} \to \mathscr{F}$ and $g\colon \mathscr{S} \to \mathscr{G}$, there exists a surjective function $h\colon \mathscr{S} \to \mathscr{H}$ with the following properties:

(i) $h \le f$.
(ii) $h \le g$.
(iii) If $h' \le f$ and $h' \le g$, then $h' \le h$.

We denote such a function h by $f \wedge g$. Then $f \wedge g$ is unique up to an invertible transformation. If all spaces and functions are measurable, then $\Sigma_{f \wedge g} = \Sigma_f \cap \Sigma_g$, where Σ_f denotes the σ-field in $\mathscr{S}$ generated by f, and so on.

Returning to the problem at hand, define $b\colon \mathscr{X} \to \mathscr{B}$ by $b = a \wedge k$ and let $\phi = b \circ k^{-1}$, $\psi = b \circ a^{-1}$. Then $B = b(X)$ is the maximal random variable which is determined both by Y [$B = \phi(Y)$] and by A [$B = \psi(A)$]. B contains the "information in common" between Y and A. It is clear that $\{a(x)\colon k(x) = y\} \subseteq \{a(x)\colon b(x) = \phi(y)\} = \{a\colon \psi(a) = \phi(y)\}$. If, for each $y \in \mathscr{Y}$, these sets are identical, we shall say that the functions $a(\cdot)$ and $k(\cdot)$ *permute*. We note, without proof, that this property is logically equivalent to any of the following, which may be easier to check in applications.

(a) There exists a surjective function $f\colon \mathscr{X} \to \mathscr{H}$ such that $f \le a$, $f \le k$, and, for all $h \in \mathscr{H}$, $\{(a(x), k(x))\colon f(x) = h\}$ is a product set.

(b) There exists f as in (a) and a function $g\colon \mathscr{Y} \to \mathscr{H}$, such that $\{a(x)\colon k(x) = y\} = \{a(x)\colon f(x) = g(y)\}$, all $y \in \mathscr{Y}$.

(c) Whenever $a(x) = a(z)$ and $k(x') = k(z)$, $x, x', z \in \mathscr{X}$, there exists $w \in \mathscr{X}$ such that $a(x') = a(w)$ and $k(x) = k(w)$.

If (a) or (b) hold, we may further deduce that $f = b$; and then, in (b), that $g = \phi$. Note that it follows from (a) that the permutation relation is in fact symmetric as between the functions a and k. At a purely formal level, the relation shares many of the abstract properties of probabilistic conditional independence [Dawid (1979)].

THEOREM 3.2. *M_2 is partitionable if and only if a and k permute. In that case the functional ancillary in M_2 may be taken to be $B = \phi(Y)$.*

PROOF. See the Appendix.

COROLLARY 3.2. *M_2 is simple if and only if the range of $(a(\cdot), k(\cdot))$ is $\mathscr{A} \times \mathscr{Y}$ [that is, the functions $a(\cdot)$ and $k(\cdot)$ are "variation-independent"].*

To summarise, we have shown that M_2 is a regular contraction of a RFM M_1 if and only if the functions $a(\cdot)$ and $k(\cdot)$ are transverse and permute. We shall then say that $a(\cdot)$ and $k(\cdot)$ *communicate*. In that case the functional ancillary in M_2 is a reduction of that in M_1 (and is, in fact, the maximal reduction of Y having this property). We remark that these conditions involve the specific form of the functional model M_1 only through the implied specification for its functional ancillary a.

EXAMPLE 3.3. We reanalyse Example 3.1 using the above conditions. We have $k(x) = \sum_{i=1}^n x_i^2$, $a(x) = (x_2 - x_1, \dots, x_n - x_1)$. The two distinct values $x = (1, 0, \dots, 0)$ and $x' = (1 - 2/n, -2/n, \dots, -2/n)$ satisfy $k(x) = k(x') = 1$, $a(x) = a(x') = (-1, -1, \dots, -1)$. Thus a and k are not transverse, and so M_2 is not invertible.

It is easy to see that $a \wedge k$ is trivial, so that M_2 will be partitionable if and only if, for all y, $\{(x_j - x_1: j = 2, \dots, n): \sum_{j=1}^n x_j^2 = y\} = \mathbb{R}^{n-1}$. But this is clearly false, since $(a_j: j = 2, \dots, n)$ is in this set if and only if

$$\sum_{j=1}^{n} (a_j - \bar{a}_n)^2 \le y$$

(defining $a_1 = 0$ and $\bar{a}_n = n^{-1}\sum_1^n a_j$). Alternatively, we can apply condition (c) with say $x = (3, 4, 0, \dots, 0)$, $x' = (5, -6, 2, \dots, 2)$ and $z = (5, 6, 2, \dots, 2)$, and note that if $a(w) = a(x') = (-11, -3, \dots, -3)$ and $k(w) = k(x) = 25$, then $25 \ge w_1^2 + (w_1 - 11)^2$, which cannot hold for any $w_1 \in \mathbb{R}$.

The regularity of M_2 in Example 3.2 may similarly be confirmed using Theorems 3.1 and 3.2.

4. Decomposition of a regular functional model. Let M_1: $\langle X = \Theta E \rangle$ be a RFM, with functional ancillary $A = a(X) = u(E)$. Let k: $\mathscr{X} \to \mathscr{Y}$ define a contraction M_2: $\langle Y = \Theta * E \rangle$, with $Y = k(X)$, $\theta * e = k(\theta e)$. Introduce the mapping m: $\mathscr{X} \to \mathscr{Y} \times \mathscr{A}$ defined as $m(\cdot) = k(\cdot) \times a(\cdot)$, so that $m(x) = (k(x), a(x))$.

THEOREM 4.1. *M_2 is a basic contraction of M_1 if and only if m is a bijection.*

PROOF. The proof follows from Theorem 3.1 and Corollary 3.2. □

It follows that, given any basic contraction M_2: $\langle Y = \Theta * E \rangle$ of M_1, we can reexpress M_1 in the form

$$M_1' : \langle (Y, A) = (\Theta * E, u(E)) \rangle \tag{4.1}$$

and then, for each θ, $\theta * E$ and $u(E)$ are variation independent. Furthermore, if $\langle Z = \Theta \circ E \rangle$ is any other basic contraction of M_1, then $Z = g(Y, A)$ where, for any value a of A, the function $g(\cdot, a)$: $\mathscr{Y} \to Z$ is a bijection.

Recasting any RFM M_1 in the form (4.1), using a basic contraction, enables us to solve for $\Theta = Y * E^{-1}$ for *any* values of Y and E, without worrying about compatibility restrictions. Fiducial inference then follows by conditioning E on $u(E) = a(x)$.

If the full model M_1 is *pivotal*, then E is determined by (Y, Θ, A). Then, for each value a of A, with E ranging over $\{e: u(e) = a\}$, the reduced model M_2 is pivotal; and conversely. A special case of this arises when there exists a pivotal basic contraction of the form $\langle Y = \Theta \circ U \rangle$, with U a function of E. The bijective correspondence $y = \theta' \circ u$ ($\theta' \in \boldsymbol{\Theta}$ fixed) then establishes that U is variation independent of A. In this case M_1 may be termed *strongly pivotal*.

EXAMPLE 4.1. Take $m = 1$ in Example 3.2, yielding the contraction M_2': $\langle Y = (\Theta + E_1)/s_E \rangle$, where $Y = X_1/s_X$. We thus have $k(x) = x_1/s_X$, $a(x) = (x_2 - x_1, \dots, x_n - x_1)$. Since M_2' is basic, $m(\cdot) = k(\cdot) \times a(\cdot)$ is a bijection. The (pivotal) model M_1 is thus equivalent to

$$M_1': \langle (X_1/s_X; X_2 - X_1, \dots, X_n - X_1) = ((\Theta + E_1)/s_E; E_2 - E_1, \dots, E_n - E_1) \rangle.$$

Fiducial inference for Θ, based on data x, is now obtainable by solving for Θ using M_2', thus expressing $\Theta = (x_1/s_x)s_E - E_1$; and assigning to E_1 its distribution conditioned on $(E_2 - E_1, \dots, E_n - E_1) = (x_2 - x_1, \dots, x_n - x_1)$. Under this condition, $s_E = s_x$ is fixed, and M_2' becomes pivotal. The fiducial model for Θ thus becomes identical with that based on the alternative pivotal contraction $\langle X_1 = \Theta + E_1 \rangle$—whose existence shows M_1 to be strongly pivotal. Yet another pivotal reduction which might be used is $\langle \bar{X} = \Theta + \bar{E} \rangle$.

In order to apply the above method, we must first construct a basic contraction. By Theorem 4.1, this involves finding a function k: $\mathscr{X} \to \mathscr{Y}$ which is *complementary* to the functional ancillary a: $\mathscr{X} \to \mathscr{A}$ in the sense that the map $k(\cdot) \times a(\cdot)$ is a bijection. This may be achieved as follows.

Choose a to be the canonical ancillary a^*, so that $a^*(x) = \mathscr{E}_x$, and $\mathscr{A}^*$ is a partition of $\mathscr{E}$. For each value α of a^*, select a representative value $z_\alpha \in \alpha$. Thus $\mathscr{Z} = \{z_\alpha: \alpha \in \mathscr{A}^*\} \subseteq \mathscr{E}$ is a *cross-section* for the partition $\mathscr{A}^*$. Since x and $z_{a^*(x)}$ are compatible and M_1 is invertible, there exists a unique $\eta(x) \in \boldsymbol{\Theta}$ with $x = \eta(x) z_{a^*(x)}$. Define $m(x) = (\eta(x), a^*(x))$. Then the map m: $\mathscr{X} \to \boldsymbol{\Theta} \times \mathscr{A}$ is clearly a bijection, and hence η: $\mathscr{X} \to \boldsymbol{\Theta}$ defines a basic contraction of M_1.

5. Fiducial prediction from a regular contraction. Let M_2: $\langle Y = \Theta * E \rangle$ be a regular contraction of the RFM M_1: $\langle X = \Theta E \rangle$, defined by k: $\mathscr{X} \to \mathscr{Y}$. Suppose that $Y = k(X)$ is observed, and we require predictive inference about X. In this section we shall attempt to apply fiducial logic to this problem.

Let $A = a(X) = u(E)$ and $B = \phi(Y) = v(E)$ be the functional ancillaries in M_1 and M_2, respectively. We know from Section 3 that B is a function of A: $B = \psi(A)$. If we now observe $Y = y$, we learn $\psi(A) = \phi(y)$. Moreover, since B is the maximal function of A observable in M_2, we can argue that this is the full extent of the information about A obtained by learning $Y = y$—this point is considered further below. This then suggests that it is now appropriate to assign to A, as an inferential distribution, its distribution conditioned on $\psi(A) = \phi(y)$. But, since M_2 is invertible, Theorem 3.1 ensures that X is completely determined by (Y, A). Thus, fixing $Y = y$ and regarding A as having its conditional distribution given $\psi(A) = \phi(y)$, we obtain an induced *fiducial predictive distribution* for X.

We remark that the above construction does not depend on the full structure of the model M_1. All we require is specification of a functional ancillary a and of a distribution for $A = a(X)$—what may be termed a "partial probabilistic model" for X. We can then make fiducial predictions for X as above, on observing $Y = k(X)$, whenever k communicates with a. While it is debatable whether "fiducial logic" should be regarded as sanctioning predictive inferences given only this minimal structure, the generality of this approach is appealing.

Returning to the case that M_1 is given as a full RFM, we can go on to construct a joint fiducial distribution for (Θ, X). Since $\Theta = Y * E^{-1}$ and $X = \Theta E^{-1}$, Θ and X are jointly determined by (Y, E). Again, we can argue that the relevant distribution to assign to E, after observing $Y = y$, is that conditional on $v(E) = \phi(y)$, and this specification, together with the fixing of Y at y, then induces a joint distribution for (Θ, X). It is clear that the margins of this distribution yield the ordinary fiducial distribution for Θ and the above predictive fiducial distribution for X, each based on data $Y = y$. It may also be verified that if we construct, from the joint fiducial distribution, the conditional distribution of Θ given $X = x$, we recover the fiducial distribution of Θ based on $X = x$.

It should, however, be noted that the conditional distribution of X given $\Theta = \theta$, in this joint fiducial distribution based on data $Y = y$, need not coincide with its sampling counterpart, viz. the conditional distribution of X given $Y = y$, calculated from the sampling distribution of X specified by $\Theta = \theta$. The distribution of E used in the former case is that conditional on $(v(E) = \phi(y), y * E^{-1} = \theta)$; and in the latter is that conditional on $(v(E) = \phi(y), \theta * E = y)$. That these are typically different is an instance of the celebrated "Borel paradox." The logical content is the same for both conditions, but they are embedded in different partitions, which can induce different conditional distributions: cf. Section 6 of Dawid and Stone (1982), from which it may be deduced that these two answers *will* coincide if the model M_2: $\langle Y = \Theta * E \rangle$ is pivotal.

EXAMPLE 5.1. Again consider Example 3.2. We have $A = (X_i - X_1: i = 2, \dots, n) = (E_i - E_1: i = 2, \dots, n)$, $B = (Y_i - Y_1: i = 2, \dots, m) = ((E_i - E_1)/s_E: i = 2, \dots, m)$ (trivial if $m = 1$). Note also that $B = (A_i/s_A: i = 2, \dots, m)$. On observing $Y = y$, we should thus regard the appropriate distribution of E, or of A, as that obtained by conditioning (if $m > 1$) on $(E_i - E_1)/s_E = y_i - y_1$, $i = 2, \ldots, m$, or on $A_i/s_A = y_i - y_1$, $i = 2, \ldots, m$. We have also learned that $X_i/s_X = y_i$, $i = 1, \ldots, m$. Given this information, and $A = (X_i - X_1 \colon i = 2, \ldots, n)$, we have $s_X = s_A$, and can thus deduce $X_1 = y_1 s_A$, whence $X_i = y_1 s_A + A_i$, $i > 1$—hence determining X as a function of A when $Y = y$ is known. (Note that other formulae are available to express this solution for X. In particular, we must have $X_i = y_i s_A$ for any $i \le m$. However, these formulae must all give identical values for X when A is constrained to satisfy $A_i/s_A = y_i - y_1$, $i = 2, \ldots, m$, as holds here as a consequence of the conditioning employed.)

The fiducial predictive distribution of X given $Y = y$ is thus that induced by expressing $X_1 = y_1 s_A$, $X_i = X_1 + A_i = y_1 s_A + A_i$, $i > 1$, where A is assigned its distribution conditioned on $A_i/s_A = y_i - y_1$, $i = 2, \ldots, m$. A joint fiducial distribution of (X, Θ) is produced by appending the equation $\Theta = y_1 s_E - E_1$ and by reexpressing the condition as $(E_i - E_1)/s_E = y_i - y_1$, $i = 2, \ldots, m$, thus extending its force to the whole of E. (Again, other formulae for Θ are available, but the conditioning renders them all equivalent.)

Conditioning further on $X = x$ in this joint distribution is equivalent to an overall conditioning of E on $\{(E_i - E_1)/s_E = y_i - y_1, i = 2, \ldots, m; s_E = x_1/y_1;$ and $E_i - E_1 = x_i - x_1, i = 2, \ldots, n\}$. These equations are mutually consistent when $y_i = x_i/s_x$, $i = 2, \ldots, m$, viz. when x is still a possible value of X after observing $Y = y$; and the above conditioning is then equivalent to conditioning just on $E_i - E_1 = x_i - x_1$, $i = 2, \ldots, n$. The consequent distribution for Θ is then just its fiducial distribution based on $X = x$.

Finally, note that, in the joint fiducial distribution, the conditional distribution of X given $\Theta = \theta$ is obtained by expressing $X_i = \theta + E_i$, and assigning to E its distribution conditioned on $(E_i - E_1)/s_E = y_i - y_1$, $i = 2, \ldots, m$, and $y_1 s_E - E_1 = \theta$. In the sampling distribution with $\Theta = \theta$, $X_i = \theta + E_i$, $Y_i = (\theta + E_i)/s_E$, so the conditional sampling distribution of X given $Y = y$ involves conditioning on $(\theta + E_i)/s_E = y_i$, $i = 1, \ldots, m$, or, equivalently, on $(E_i - E_1)/s_E = y_i - y_1$, $i = 2, \ldots, m$, and $(E_1 + \theta)/s_E = y_1$. Although $y_1 s_E - E_1 = \theta$ holds if and only if $(E_1 + \theta)/s_E = y_1$, conditioning on the value θ for $y_1 s_E - E_1$ is quite different from conditioning on the value y_1 for $(E_1 + \theta)/s_E$, so that the two ways of constructing a distribution for X given (Θ, Y) will not agree. Indeed, by changing variables it may readily be shown that the sampling density for X_1 given $Y = y$, when $\Theta = \theta$, is proportional to $|x_1| f(x_1 - \theta, x_1/y_1)$, where $f(e, s)$ is the joint density function of (E_1, s_E) conditional on $(E_i - E_1)/s_E = y_i - y_1$, $i = 2, \ldots, m$; while the predictive fiducial density for X_1 given $\Theta = \theta$, based on data $Y = y$, is proportional to $f(x_1 - \theta, x_1/y_1)$.

Finally in this section, we reconsider the claim that the only information about A [or E], obtained by observing $Y = y$, is that $\psi(A) = \phi(Y)$ [or $v(E) = \phi(Y)$]. Let M_3: $\langle Z = \Theta \circ E \rangle$ be a basic contraction of M_2, shown to exist in Section 4. Theorem 4.1 shows that we can reexpress Y as (Z, B), and thus M_2 as M_2^*: $\langle (Z, B) = (\Theta \circ E, v(E)) \rangle$. On learning $Y = y$, or, equivalently, $(Z = z, B = b)$, we know (i) $v(E) = b$ and (ii) $\Theta \circ E = z$. However, since the model M_3 is basic, any value of E is compatible with (ii), for some value θ of Θ. The same logic as governs ordinary fiducial logic now holds that, if Θ is initially completely unknown, no useful information about E can be extracted from (ii), and hence that the condition $v(E) = b$ is the full extent of the information gained about E.

6. Chainable RFMs. In this section we lay the algebraic groundwork for an extension of fiducial prediction to cases in which the observed Y is not a function of the predictand X. All proofs are confined to the Appendix.

Suppose then that, in addition to the functional model M_1: $\langle X = \Theta E \rangle$, a RFM with functional ancillary $A = a(X) = u(E)$, we have another functional model M_2: $\langle Y = X \circ F \rangle$ (not necessarily regular) relating Y to X.

Note that, in M_2, we are taking the "parameter" to be X, so that M_2 is a model for the structure of Y, conditional on the value of X. We regard (E, F) as logically independent, with a joint distribution. In most cases of interest, E and F will be probabilistically independent. Then the model M_2 for Y given X is unaffected by further (exact or probabilistic) information about Θ in M_1, and the sequence $\Theta \to X \to Y$ is a (functional) Markov chain. In particular, in the model for (X, Y) given Θ, X is sufficient.

As an example of this structure, we might have a model in which X and Y are sufficient statistics based on samples of size N and n, $N > n$, respectively, where we require predictive inferences for X after observing Y; and functional models are used to express both the sampling distributions for X given Θ and the conditional distributions for Y given X (which will not involve Θ). Direct specification of these conditional distributions lies at the heart of the theory of *extreme-point modelling* [Lauritzen (1982)], and the present approach can indeed be used to construct and analyse a functional-model analogue of that theory, although we shall not do so here. As another example, X might represent a *parameter* of the distribution of Y, so that M_2 is an ordinary parametric functional model, while M_1 describes a partial prior probability structure for this parameter, in terms of a "hyper-parameter" Θ. Then we would wish to draw "predictive" inference about X based on observation of Y, taking due account of the prior structure. We explore this interpretation further in Section 8.

The models M_1 and M_2 may be compounded to produce

$$M_3 : \langle Y = \Theta * (E, F) \rangle, \quad \text{where } \theta * (e, f) = (\theta e) \circ f.$$

M_3 describes the "sampling model" for Y as governed by the parameter Θ. If it is regular, we can use it to construct fiducial inferences for Θ—and, as we shall see in Section 7, fiducial predictions for X—based on the observation Y. In this case we shall call M_1 and M_2 *chainable*. We now investigate this property.

We first note that M_3 is a contraction of the regular model

$$M_0 : \langle (X, S) = \Theta \otimes (E, F) \rangle,$$

where $\Theta \otimes (E, F) = (\Theta E, F)$, and the contraction is effected by the function

$(x, s) \mapsto x \circ s$. The functional ancillary in M_0 is given by the function $(x, s) \mapsto (a(x), s)$. The following is then immediate from Theorem 3.1.

THEOREM 6.1. *M_3 is invertible if and only if, for any f,*

$$\left.\begin{array}{r} a(x_1) = a(x_2) \\ x_1 \circ f = x_2 \circ f \end{array}\right\} \Rightarrow x_1 = x_2.$$

Note that this condition may be restated: "For each f, the functions $a(\cdot)$ and $k_f(\cdot)$ are transverse," where $k_f(x) = x \circ f$.

THEOREM 6.2. *M_3 is partitionable if and only if, for each f, the contraction M_f of M_1 defined by k_f is partitionable, and there exists a function $c\colon \mathscr{Y} \to \mathscr{K}$ which serves as a functional ancillary in each M_f. In this case $C = c(Y)$ is the functional ancillary in M_3.*

PROOF. See the Appendix.

COROLLARY 6.2. *M_3 is simple if and only if each M_f is simple.*

To summarize, we have shown that M_1 and M_2 are chainable if and only if, for each f, the mapping $x \mapsto x \circ f$ defines a regular contraction M_f of M_1, with a common functional ancillary in every M_f, which then serves as the functional ancillary in M_3.

EXAMPLE 6.1. Take $\Theta = \mathbb{R}^1$, $\mathscr{X} = \mathscr{Y} = \mathscr{E} = \mathbb{R}^2$, $\mathscr{F} = \mathbb{R}^1 \times \mathbb{R}^+$, M_1: $\langle (X_1, X_2) = (\Theta + E_1, \Theta + E_2) \rangle$ and M_2: $\langle (Y_1, Y_2) = ((X_1 + F_1)/F_2, (X_2 + F_1)/F_2) \rangle$. Then we obtain M_3:$\langle (Y_1, Y_2) = ((\Theta + E_1 + F_1)/F_2, (\Theta + E_2 + F_1)/F_2) \rangle$ and M_f: $\langle (Y_1, Y_2) = ((\Theta + E_1 + f_1)/f_2, (\Theta + E_2 + f_1)/f_2) \rangle$. We note that M_1 is regular, with functional ancillary $A = X_1 - X_2 = E_1 - E_2$; and that M_2 is basic. For fixed f, M_f is isomorphic to M_1, and hence is a regular contraction. The functional ancillary in M_f can be taken as $Y_1 - Y_2 = (E_1 - E_2)/f_2$, the same (as a function of Y) for all $f \in \mathscr{F}$. We deduce that M_1 and M_2 are chainable, and thus M_3 is regular, with functional ancillary $Y_1 - Y_2 = (E_1 - E_2)/F_2$—as indeed is readily apparent.

7. Inference and prediction in chained models. Let M_1 and M_2 be chainable models, as in Section 6, with composition M_3. We shall assume henceforth that E and F are independent.

We can again regard M_3, with functional ancillary $C = c(Y) = w(E, F)$ say, as the regular contraction $Y = X \circ S$ of the RFM M_0^*: $\langle (X, S) = \Theta \otimes (E, F) \rangle$. If we now observe $Y = y$, we can thus make fiducial prediction for (X, S) jointly, as in Section 5, and then marginalize to obtain a fiducial predictive distribution for X based on Y. Essentially, we express X as a function of Y, E and F, and then use the observation both to fix Y and to update the joint distribution of (E, F) by conditioning on $w(E, F) = c = c(y)$, the observed value of C. We shall call this method the "global" approach. We can also apply the same approach to yield a fiducial distribution for Θ, or for (Θ, X) jointly. In all cases the relevant distribution for (E, F) is obtained by conditioning on $w(E, F) = c$.

An alternative approach is possible when M_2 is also a RFM, with functional ancillary $B = b(Y) = v(F)$ say. We can first concentrate on M_2 to obtain a fiducial distribution for its "parameter" X, based on observing $Y = y$. Then, for each value of x, we can use M_1 alone to obtain a fiducial distribution for Θ given $X = x$. The Markov property of the original chain suggests that this may likewise be taken as the relevant fiducial distribution for Θ given both $X = x$ and $Y = y$. We can then combine these two ingredients—a fiducial distribution for X and a family of fiducial conditional distributions for Θ given X—into a joint "stepwise" fiducial distribution for (Θ, X) given $Y = y$.

We now investigate the relationships between these two approaches when M_2 is a RFM. The global approach provides directly a joint inferential distribution for (E, F)—that produced by conditioning on $w(E, F) = c$. This in turn induces the joint inferential distribution for $X = y \circ F^{-1}$ and $\Theta = (y \circ F^{-1})E^{-1}$. We can then consider this as decomposed, by conditioning, into a marginal fiducial predictive distribution of X, and conditional fiducial distributions for Θ, given X. The stepwise approach, on the other hand, supplies each of these components directly. We shall compare the corresponding components of this decomposition, as produced by the two approaches.

THEOREM 7.1. *The conditional fiducial distributions for Θ given X are the same in both the global and the stepwise approaches.*

PROOF. Consider forming, in the global fiducial distribution, the conditional distribution for Θ, given $X = x$. (Of course, we must require that $X = x$ be compatible in M_2 with the observation $Y = y$, so that, for some $f \in \mathscr{F}$, $y = x \circ f$—and thus $x = y \circ f^{-1}$.) This is the distribution of xE^{-1}, given $y \circ F^{-1} = x$, as calculated from the joint distribution of (E, F) already conditioned on $w(E, F) = c$. We thus need the distribution of E, calculated by conditioning on both $y \circ F^{-1} = x$ and $w(E, F) = c$ in the original joint distribution of (E, F). Now note that $w(e, f) = c(x \circ f)$ if and only if, for some θ, $x \circ f = (\theta e) \circ f$ which, since M_2 is invertible, holds if and only if $x = \theta e$ for some θ, or, equivalently, $a(x) = u(e)$. When $y = x \circ f$, we thus have $w(e, f) = c(y)$ if and only if $a(x) = u(e)$. So the conditions "$y \circ f^{-1} = x$ and $w(e, f) = c(y)$" are together equivalent to "$y \circ f^{-1} = x$ and $u(e) = a(x)$." Hence the desired conditional fiducial distribution for Θ given $X = x$ is, equivalently, obtained from the representation $\Theta = xE^{-1}$ by assigning to E its original distribution conditioned on $y \circ F^{-1} = x$ and $u(E) = a(x)$. Since initially E and F are independent, this in turn is equivalent to conditioning only on $u(E) = a(x)$. But this is exactly the prescription delivered by the stepwise approach for inference about Θ, given $X = x$. (Note that this proof still works if M_2 is invertible, but not partitionable.) □

When we turn to the marginal fiducial distribution for X, however, the two approaches will generally disagree. Each expresses $X = y \circ F^{-1}$, but whereas the stepwise approach uses the distribution of F conditioned on $v(F) = b(y)$, the global approach conditions on $w(E, F) = c(y)$ in the joint distribution.

We may note that $B = b(Y)$ is a function of $C = c(Y)$ [and thus $v(F)$ is the same function of $w(E, F)$]. For $c(y_1) = c(y_2) \Rightarrow$ there exist $\theta_1, \theta_2 \in \mathbf{\Theta}$ and (e, f) satisfying $w(e, f) = c(y_1) = c(y_2)$, such that $y_1 = (\theta_1 e) \circ f$ and $y_2 = (\theta_2 e) \circ f$. In particular, $y_1 = x_1 \circ f$, $y_2 = x_2 \circ f$ (with $x_1 = \theta_1 e$, $x_2 = \theta_2 e$), so that $b(y_1) = v(f) = b(y_2)$.

We thus see that the global approach generally involves conditioning on additional information, over and above that of the stepwise approach, and so the two approaches generally produce different fiducial distributions for X, and hence for Θ.

We again remark that the conditional distribution of X given $\Theta = \theta$ in the fiducial distribution given $Y = y$ (whether global or stepwise) need not coincide with its sampling counterpart, the conditional distribution of X given $Y = y$, when $\Theta = \theta$. However, if the model M_3: $\langle Y = \Theta * (E, F) \rangle$ is pivotal, then these will agree if we use the global approach.

EXAMPLE 7.1. Take M_1 and M_2 as in Example 6.1. The global predictive distribution for X, based on $Y = y$, is obtained from $X_1 = y_1 F_2 - F_1$, $X_2 = y_2 F_2 - F_1$ by conditioning (F_1, F_2) on the ancillary information in M_3, namely $(E_1 - E_2)/F_2 = y_1 - y_2$, in the joint distribution of (E_1, E_2, F_1, F_2). Since M_2 is basic, the stepwise predictive distribution of X involves no conditioning at all. And since, even with (F_1, F_2) independent of (E_1, E_2), (F_1, F_2) will not generally be independent of $(E_1 - E_2)/F_2$, these two results will disagree.

The global conditional distribution for Θ given $X = x$ is obtained from $\Theta = x_1 - E_1$ by conditioning E on $(E_1 - E_2)/F_2 = y_1 - y_2$, $y_1 F_2 - F_1 = x_1$ and $y_2 F_2 - F_1 = x_2$. These equations are equivalent to $F_1 = (y_1 x_2 - x_1 y_2)/(y_2 - y_1)$, $F_2 = (x_2 - x_1)/(y_2 - y_1)$ and $E_1 - E_2 = x_1 - x_2$. By independence, we thus need only condition on $E_1 - E_2 = x_1 - x_2$, exactly as for the stepwise approach.

The global fiducial conditional distribution for X given $\Theta = \theta$, based on data $Y = y$, is obtained by expressing $X_i = \theta + E_i$ and conditioning on $(E_1 - E_2)/F_2 = y_1 - y_2$ and $y_1 F_2 - E_1 - F_1 = \theta$. The sampling counterpart of this distribution would involve conditioning on $(E_1 - E_2)/F_2 = y_1 - y_2$ and $(\theta + E_1 + F_1)/F_2 = y_1$. Because these logically equivalent conditions involve different partitions, the results will generally disagree.

One case in which the two approaches yield identical fiducial predictive distributions for X is when M_3 is simple, since then C, and hence also B, is trivial, and no conditioning is required in either approach. By Corollary 6.2 this is the case if and only if each M_f is simple.

EXAMPLE 7.2. Take $\mathscr{X} = \mathscr{Y} = \mathbf{\Theta} = \mathbb{R}^1$, and suppose that each of the models M_1: $\langle X = \Theta E \rangle$ and M_2: $\langle Y = X \circ F \rangle$ is invertible and *monotone* [Dawid and

Stone (1982), subsection 4.2], so that, for any $e \in \mathscr{E}$, $f \in \mathscr{F}$, $\theta_1 \leq \theta_2 \Leftrightarrow \theta_1 e \leq \theta_2 e$, and $x_1 \leq x_2 \Leftrightarrow x_1 \circ f \leq x_2 \circ f$. It follows that M_3: $\langle Y = \Theta * (E, F) \rangle$ is likewise monotone, and all three models are basic.

Let $F_\theta(\cdot)$ be the distribution function for X given $\Theta = \theta$ implied by M_1, $G_x(\cdot)$ that for Y given $X = x$ implied by M_2, and $H_\theta(\cdot)$ that for Y given $\Theta = \theta$ implied by M_3. Then

$$H_\theta(y) = \int G_x(y)\, dF_\theta(x). \tag{7.1}$$

By Lemma 4.1 of Dawid and Stone (1982), the global fiducial distribution Π_y for Θ given $Y = y$ satisfies

$$\Pi_y(\Theta \geq \theta) = H_\theta(y). \tag{7.2}$$

Similarly, the stepwise approach delivers the fiducial distributions Γ_y for X given $Y = y$ and Φ_x for Θ given $X = x$, where

$$\Gamma_y(X \geq x) = G_x(y) \tag{7.3}$$

and

$$\Phi_x(\Theta \geq \theta) = F_\theta(x). \tag{7.4}$$

The implied stepwise distribution Π'_y for Θ given $Y = y$ then satisfies

$$\begin{aligned} \Pi'_y(\Theta \geq \theta) &= -\int \Phi_x(\Theta \geq \theta)\, d\Gamma_y(X \geq x) \\ &= -\int F_\theta(x)\, d_x G_x(y). \end{aligned} \tag{7.5}$$

Since M_3 is basic, however, Π_y and Π'_y must agree. This implies that

$$H_\theta(y) = -\int F_\theta(x)\, d_x G_x(y), \tag{7.6}$$

as indeed follows from (7.1) by partial integration (under boundary conditions which are satisfied in the present context).

Note that the exact functional structure of the component monotone models is irrelevant. Whenever the families $\{F_\theta: \theta \in \boldsymbol{\Theta}\}$ and $\{G_x: x \in \mathscr{X}\}$ are *Fisherian* [see Example 2.1 of Dawid and Stone (1982)], so is $\{H_\theta: \theta \in \boldsymbol{\Theta}\}$, and the fiducial distributions implied by any underlying monotone models are just those produced by Fisher's technique [Fisher (1930)] based on the probability integral transform. We have thus shown that inference in a chain of such models is the same, whether performed globally or step by step. (Note, however, that the fiducial and sampling distributions for X given Θ and Y need not agree.)

When M_3 is not basic, we have the following result.

THEOREM 7.2. *If $w(E, F)$ is expressible in the form $k(E, v(F))$, then the global and stepwise approaches will yield the same fiducial distribution for X.*

PROOF. Since E and F are independent, they remain so conditional on $v(F) = v$. In this conditional distribution, $w(E, F) = k(E, v(F))$ may be replaced by $k(E, v)$, a function of E alone, and is thus independent of F. We have thus shown that $w(E, F)$ and F are independent conditional on $v(F)$, so that the conditional distribution of F given $(v(F), w(E, F))$ is the same as that given $v(F)$ alone. But $v(F)$ is a reduction of $w(E, F)$, so that the former conditioning is just on $w(E, F)$, as required for stepwise analysis; while the latter is that appropriate for global analysis. □

It does not appear easy to reexpress the condition of Theorem 7.2 simply in terms of the structure of the constituent functional models. Note that $w(E, F)$ is always a reduction of $(u(E), F)$, so that the condition is equivalent to being able to express the ancillary $C = w(E, F)$ in M_3 as a function of the ancillaries $A = u(E)$ in M_1 and $B = v(F)$ in M_2.

Which approach? We have seen that, when we have chainable RFMs M_1: $\langle X = \Theta E \rangle$ and M_2: $\langle Y = X \circ F \rangle$, the global and stepwise approaches will generally deliver different fiducial predictive distributions for X, given data $Y = y$. Which (if either) should be preferred? Recall that the logic underlying fiducial inference is only supposed to apply when, initially, "nothing is known" about the parameter. The stepwise approach uses the fiducial method on M_2 to infer about its "parameter" X from $Y = y$. However, it is not true that initially nothing is known about X. For instance, if M_1 imposes $X_i = \Theta \times E_i$, $i = 1, 2$, then, even when nothing is known about Θ, we do have a fully specified distribution for X_1/X_2, so that it is plainly wrong to assert that nothing is known about X. Consequently, we argue that the stepwise approach, tempting though it is by virtue of its simplicity, is incorrect. The global approach, which conditions on additional information, produces the correct fiducial distributions—so long, at any rate, as we have no further information about the parameter Θ of M_1, but can regard that as initially "completely unknown."

The considerations of Sections 3 and 5 through 7 may be extended to the case that we have a sequence of more than two models, suitably related by contraction, or by "single-step" functional models.

8. Bayesian and semi-Bayesian inference. In this section we investigate connections between Bayesian and fiducial inference and a compromise between them. The idea of such a common framework has been promoted by Barnard (1985).

Consider first the trivial functional model M_1: $\langle X = E \rangle$, with effectively a one-point parameter space; together with an invertible functional model M_2: $\langle Y = X \circ F \rangle$, with E and F independent. Their composition is M_3: $\langle Y = E \circ F \rangle$, and both M_1 and M_3 are trivially regular, with respective functional ancillaries $A = X = E$ and $C = Y = E \circ F$. On observing $Y = y$, the fiducial predictive distribution for $X = E$ is thus obtained by conditioning on $E \circ F = y$, viz. on $Y = y$. In other words, we simply condition the original (known) joint distribution of (X, Y) on $Y = y$. In this degenerate case, fiducial inference thus agrees (fortunately!) with the standard prescriptions of probability theory.

Now consider an arbitrary RFM $\langle X = \Theta E \rangle$. In the absence of any prior information about Θ, we should use the fiducial distribution of Θ to make inference from data $X = x$.

However, suppose instead that we have a known prior distribution Π for Θ. This can be expressed in the form of a trivial functional model $\langle \Theta = F, F \sim \Pi \rangle$; moreover, since it is implicit in any functional model $\langle X = \Theta E \rangle$ that the distribution of E is the same for all values of Θ, and here $\Theta = F$, we have E and F independent. The argument above now applies to yield, as the relevant fiducial distribution for Θ based on data $X = x$, that calculated by conditioning on $X = x$ in the induced joint distribution for (Θ, X). In other words, in the presence of a fully specified prior distribution, fiducial inference is the same as Bayesian inference.

The above cases—no knowledge of Θ and a fully specified prior distribution for Θ—are just the two extremes of a range of possibilities. More generally, we could have partial prior information about Θ, specified in terms of a functional model $\langle \Theta = \Lambda \circ F \rangle$, where the hyper-parameter Λ is completely unknown (alternatively, this itself could be just the first link in a chain of models constituting a hierarchical structure for Θ). The techniques of Section 7 can then be used to derive the fiducial distribution of Θ relevant for the specific degree of knowledge implicit in such structuring.

EXAMPLE 8.1. Let $\mathscr{X} = \mathbf{\Theta} = \mathscr{E} = \mathbb{R}^n$, with the sampling model embodied in M_1: $\langle X_i = \Theta_i + E_i,\ i = 1, \dots, n,\ (E_i) \sim NID(0, 1) \rangle$. Thus $(X_i) \sim NID(\Theta_i, 1)$. We consider inference for Θ based on $X = x$, in various states of prior information about Θ.

(i) *No prior information*. The appropriate inferential distribution is just the fiducial distribution constructed from M_1. Since M_1 is simple, this yields $\Theta_i = x_i - E_i$ with $(\Theta_i) \sim NID(x_i, 1)$. This happens to agree with the formal Bayesian posterior derived from a uniform prior for Θ.

(ii) *Full prior information*. Suppose, for example, $(\Theta_i) \sim NID(\lambda, 1)$, with λ known. We can express this as

$$(8.1) \qquad M_{2,\lambda}: \langle \Theta_i = \lambda + F_i, (F_i) \sim NID(0,1) \rangle.$$

The compound model for X is then

$$(8.2) \qquad M_{3,\lambda}: \langle X_i = \lambda + U_i, (U_i) \sim NID(0,2) \rangle,$$

where $U_i = E_i + F_i$. This is trivially regular, with functional ancillary $U = (U_i: i = 1, \dots, n)$ (recall that λ is known). Consequently, the relevant inferential distribution conditions E on $E_i + F_i = x_i - \lambda$, $i = 1, \dots, n$, which, with (E_i) and (F_i) all initially $NID(0, 1)$, produces $(E_i) \sim NID(\frac{1}{2}(x_i - \lambda), \frac{1}{2})$. Then $\Theta_i =$

$x_i - E_i$ yields $(\Theta_i) \sim NID(\frac{1}{2}(x_i + \lambda), \frac{1}{2})$—identical with the posterior distribution of Θ for the specified prior distribution.

(iii) *Partial prior information*. Now suppose (8.1) is replaced by

$$M_2: \langle \Theta_i = \Lambda + F_i, (F_i) \sim NID(0,1) \rangle, \tag{8.3}$$

with Λ completely unknown. The compound model is

$$M_3: \langle X_i = \Lambda + U_i, (U_i) \sim NID(0,2) \rangle. \tag{8.4}$$

This is again regular, but now the functional ancillary can be taken as $(X_i - \bar{X}: i = 1, \dots, n) = (U_i - \bar{U}: i = 1, \dots, n)$ $(\bar{X} = \sum_{i=1}^n X_i/n$, etc.). We must therefore condition on $(U_i - \bar{U}) = (x_i - \bar{x})$, $i = 1, \dots, n$. Let $V_i = E_i - F_i$. We have $\Theta_i = \frac{1}{2}(X_i + \bar{X}) - \frac{1}{2}\bar{U} - V_i$, $\Lambda = \bar{X} - \bar{U}$, with $\bar{U} \sim N(0, 2/n)$, $(V_i) \sim NID(0,2)$, independently of each other and of $(U_i - \bar{U}: i = 1, \dots, n)$. Thus the fiducial marginal distribution of Λ is $N(\bar{x}, 2/n)$, and the fiducial conditional distribution of Θ given $\Lambda = \lambda$ is $(\Theta_i) \sim NID(\frac{1}{2}(x_i + \lambda), \frac{1}{2})$, this latter agreeing with case (ii) above because the model M_3 is pivotal. These inferential distributions are the same as formal Bayes posterior distributions based on the improper hierarchical prior: $(\Theta_i)|\Lambda \sim NID(\Lambda, 1)$, Λ uniform.

Note that, if $n = 1$, both cases (i) and (iii) yield the same distribution. For data x from the model $\langle X = \Theta + E,\ E \sim N(0,1) \rangle$, inference about Θ is the same whether Θ is regarded as completely unknown or structured by $\langle \Theta = \Lambda + F,\ F \sim N(0,1) \rangle$, with Λ completely unknown.

EXAMPLE 8.2. Let $\mathscr{X} = \mathbf{\Theta} = \mathbb{R}^2$, $\mathscr{E} = \mathbb{R}^1 \times \mathbb{R}^+$, with the sampling model expressed as M_1: $\langle X_i = (\Theta_i + E_1)/E_2,\ i = 1,2 \rangle$.

(i) *No prior information*. Since M_1 is basic, the relevant inferential distribution for Θ based on $X = x$ has $\Theta_i = x_i E_2 - E_1$, with (E_1, E_2) assigned its initial distribution. Expressing $E_1 = (\Theta_1 X_2 - \Theta_2 X_1)/(X_1 - X_2)$, $E_2 = (\Theta_1 - \Theta_2)/(X_1 - X_2)$, we find $\partial(e_1, e_2)/\partial(x_1, x_2) = (\theta_1 - \theta_2)^2/(x_1 - x_2)^3 = J_1$, say, and $\partial(e_1, e_2)/\partial(\theta_1, \theta_2) = 1/(x_1 - x_2) = J_2$, say. Since the sampling density is $f(x|\theta) = f_E(e)|J_1|$ and the fiducial density is $\pi(\theta|x) = f_E(e)|J_2|$, where f_E is the unconditional density of E, we find that this fiducial solution agrees with the formal Bayes posterior based on the improper prior density $\pi(\theta_1, \theta_2) \propto (\theta_1 - \theta_2)^{-2}$.

(ii) *Partial prior information*. Now take the prior knowledge about Θ to be structured as M_2: $\langle \Theta = \Lambda + F_i,\ i = 1,2 \rangle$, with the distribution of (F_1, F_2) specified but Λ completely unknown. We obtain the compound model

$$M_3: \langle X_i = (\Lambda + F_i + E_1)/E_2, i = 1,2 \rangle.$$

This is again regular, with functional ancillary $X_1 - X_2 = (F_1 - F_2)/E_2$. The relevant fiducial distribution for Θ is thus obtained by conditioning on $(F_1 - F_2)/E_2 = x_1 - x_2$, yielding a conditional density for (E_1, E_2) proportional to $|e_2| f_E(e_1, e_2)$. A matching improper prior density for Θ is now $\pi(\theta_1, \theta_2) \propto |\theta_1 - \theta_2|^{-1}$. Similarly, the fiducial density for Λ is the Bayes posterior, based on the sampling model for $\bar{X}$ given Λ, and an improper uniform prior density for Λ. However, the joint fiducial density for $(\Lambda, \Theta_1, \Theta_2)$ is the posterior obtained from the prior $\pi(\theta_1, \theta_2, \lambda) \propto |\theta_1 - \theta_2|^{-1} f_F(\theta_1 - \lambda, \theta_2 - \lambda)$, f_F being the density for (F_1, F_2). If we had, instead, started with a *known* value $\Lambda = \lambda$, the correct analysis would have been the fully Bayesian one obtained using the implied prior density $f_F(\theta_1 - \lambda, \theta_2 - \lambda)$. Thus in this case the effect of incorporating the information "$\Lambda = \lambda$" depends on whether this is done before or after observing the data.

In the above examples, all the fiducial distributions happen to have a Bayesian interpretation; that this need not always be the case is evidenced by Example 2.2 of Dawid and Stone (1982).

9. Discussion. Our analysis has shown how, under certain algebraic restrictions, the logic of "fiducial inversion" in a functional model may be extended to problems of prediction. We have also uncovered some problematic aspects of this approach, to add to the "conditioning inconsistency" of parametric fiducial inference displayed by Dawid and Stone (1982). In particular, the predictive fiducial distribution for X based on Y, when conditioned on Θ, need not coincide with the sampling distribution of X given Θ, conditioned on Y. Not only is this a clearly unacceptable inconsistency with the sampling model, it may be transformed into an internal inconsistency of fiducial inference on noting that, were a prior distribution to be specified for Θ, fiducial inference would then coincide with ordinary (Bayesian) probabilistic analysis, leading to a conditional fiducial distribution for X given Θ, based on Y, which *is* consistent with the sampling model. It is paradoxical that a distribution conditioned on Θ should be affected by whether or not a marginal distribution for Θ is assigned. In the semi-Bayesian context of Section 8, the inferential distribution for Θ, conditioned on $\Lambda = \lambda$, is seen to depend on when this condition is incorporated—again unacceptable behaviour.

Another inconsistency has been noted, that between global and stepwise inference in chained models. However, we have argued that this presents no logical difficulty, since only the global method is properly consistent with fiducial logic.

What lessons may be drawn from our analysis? Whilst fiducial logic, in the context of regular functional models, has a fair degree of logical cohesion, it nevertheless produces some unacceptable answers. These appear in problems involving conditioning. They can, however, be made to disappear if one restricts attention to *pivotal* models, in which the error variable is required to be a function of the data and the parameter. It thus appears that fiducial logic may be more generally applicable when restricted to such pivotal models. Even though this restriction will not remove all the inconsistencies between fiducial and sampling distributions—such as the *strong inconsistency* of Stone (1976) —it may be that it is sufficient to ensure a measure of internal self-consistency. However, this requires further investigation.

APPENDIX

Proofs of algebraic results.

PROOF OF THEOREM 3.1. Assume first that M_2 is invertible, and suppose $a(x_1) = a(x_2)$ and $k(x_1) = k(x_2)$. Take $e \in \mathscr{E}$ with $u(e) = a(x_1) = a(x_2)$. Then there exist $\theta_1, \theta_2 \in \mathbf{\Theta}$ with $x_1 = \theta_1 e$ and $x_2 = \theta_2 e$. Thus $\theta_1 * e = k(x_1) = k(x_2) = \theta_2 * e$, and so $\theta_1 = \theta_2$, whence $x_1 = x_2$.

Now assume that $a(\cdot)$ and $k(\cdot)$ are transverse and that $\theta_1 * e = \theta_2 * e$. Then, with $x_1 = \theta_1 e$ and $x_2 = \theta_2 e$, $k(x_1) = k(x_2)$ and $a(x_1) = u(e) = a(x_2)$. Hence $x_1 = x_2$, whence $\theta_1 = \theta_2$ by invertibility of M_1. □

PROOF OF THEOREM 3.2. It is enough to prove the result when a is the canonical ancillary in M_1. The canonical reduction of M_2 is c: $\mathscr{Y} \to 2^{\mathscr{E}}$, where $c(y) = \{e\colon y = k(\theta e), \text{ some } \theta \in \mathbf{\Theta}\}$. Then $c(y) = \bigcup\{\mathscr{E}_x\colon k(x) = y\} = \bigcup\{a(x)\colon k(x) = y\}$. Letting $\mathscr{C}_y = \{a(x)\colon k(x) = y\}$, it is clear that the range of c is a partition of $\mathscr{E}$, and thus M_2 is partitionable, if and only if, for all $y_1, y_2 \in \mathscr{Y}$, $\mathscr{C}_{y_1}$ and $\mathscr{C}_{y_2}$ are either disjoint or identical.

If a and k permute, then $\mathscr{C}_y = \{\alpha \in \mathscr{A}\colon \psi(\alpha) = \phi(y)\}$. Thus $\mathscr{C}_{y_1} = \mathscr{C}_{y_2}$ if $\phi(y_1) = \phi(y_2)$, while $\mathscr{C}_{y_1} \cap \mathscr{C}_{y_2} = \varnothing$ otherwise, and hence M_2 is partitionable. Clearly also in this case $c(y_1) = c(y_2)$, or, equivalently, $\mathscr{C}_{y_1} = \mathscr{C}_{y_2}$, if and only if $\phi(y_1) = \phi(y_2)$, so that $B = \phi(Y)$ may be taken as functional ancillary.

Conversely, suppose M_2 is partitionable, and define $b(x) = \mathscr{C}_{k(x)}$. Clearly $b \le k$. Also $a(x) \in b(x)$, which property, by partitionability, determines $b(x)$ uniquely, whence $b \le a$. Furthermore, $a(x) \in \mathscr{C}_y$ if and only if $C_y = b(x)$, so that $\{a(x)\colon k(x) = y\} = \mathscr{C}_y = \{a(x)\colon b(x) = \mathscr{C}_y\}$. Thus, applying property (b) in Section 3, a and k permute, and $b = a \wedge k$. □

PROOF OF THEOREM 6.2. Let a^*, c^* and c_f^* be the canonical reductions in M_1, M_3 and M_f, respectively. Then $c_f^*(y) = \bigcup\{a^*(x)\colon x \circ f = y\}$, while $(e, f) \in c^*(y)$ if and only if, for some $x \in \mathscr{X}$ and $\theta \in \mathbf{\Theta}$, $\theta e = x$ and $x \circ f = y$, that is, for some x, $e \in a^*(x)$ and $x \circ f = y$. Thus $c^*(y) = \bigcup_x a^*(x) \times \{f\colon x \circ f = y\} = \bigcup_f c_f^*(y) \times \{f\}$. The range of c^* is thus a partition if and only if:

(i) The range of c_f^* is a partition, for each f.
(ii) Whenever $c_f^*(y_1) = c_f^*(y_2)$ for some f, this holds for all $f \in \mathscr{F}$.

Condition (i) states that each M_f is partitionable. Fix $f_0 \in \mathscr{F}$, and define $c(y) = c_{f_0}^*(y)$. Then condition (ii) states that, for any $f \in \mathscr{F}$, c is an invertible transformation of c_f^*, and so serves as a functional ancillary in each M_f. Clearly c is then also an invertible transformation of c^*, and hence a functional ancillary in M_3. □

(It may be noted that the above proof does not require either M_1 or M_2 to be regular.)

Acknowledgments. We wish to thank an Associate Editor and referees for their helpful comments on an earlier draft of this paper.

REFERENCES

BARNARD G. A. (1980). Pivotal inference and the Bayesian controversy (with discussion). In *Bayesian Statistics* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 295–318. Univ. Valencia Press.

BARNARD, G. A. (1985). A coherent view of statistical inference. Technical report, Dept. Statistics, Univ. Waterloo.

DAWID, A. P. (1979). Conditional independence in statistical theory (with discussion). *J. Roy. Statist. Soc. Ser. B* **41** 1–31.

DAWID, A. P. and STONE, M. (1982). The functional-model basis of fiducial inference (with discussion). *Ann. Statist.* **10** 1054–1074.

FISHER, R. A. (1930). Inverse probability. *Proc. Cambridge Philos. Soc.* **26** 528–535.

FISHER, R. A. (1935). The fiducial argument in statistical inference. *Ann. Eugenics* **6** 391–398.

FISHER, R. A. (1973). *Statistical Methods and Scientific Inference* 3rd ed. Hafner, New York.

FRASER, D. A. S. (1968). *The Structure of Inference*. Wiley, New York.

HORA, R. B. and BUEHLER, R. J. (1967). Fiducial theory and invariant prediction. *Ann. Math. Statist.* **38** 795–801.

KITAGAWA, T. (1957). Successive process of statistical inference associated with an additive family of sufficient statistics. *Bull. Math. Statist. Fukuoka* **7** 92–112.

LAURITZEN, S. L. (1982). *Statistical Models as Extremal Families*. Aalborg Univ. Press.

STONE, M. (1976). Strong inconsistency from uniform priors (with discussion). *J. Amer. Statist. Assoc.* **71** 114–125.

DEPARTMENT OF STATISTICAL SCIENCE
UNIVERSITY COLLEGE LONDON
GOWER STREET
LONDON WC1E 6BT
ENGLAND

DEPARTMENT OF MATHEMATICAL STATISTICS
EAST CHINA NORMAL UNIVERSITY
SHANGHAI 200062
CHINA