WBS Economic Modelling & Forecasting Group

Working Paper Series: No 2
July 2013

Working paper title: The Recalibrated and Copula Opinion Pools

James Mitchell
Warwick Business School
University of Warwick

The Recalibrated and Copula Opinion Pools

James Mitchell*
Warwick Business School, University of Warwick, U.K.

July 23, 2013

Abstract

This paper develops methods for combining density forecasts which accommodate stochastic dependence between different experts' predictions. Previous work combining density forecasts, using so-called "opinion pools", has essentially ignored dependence. The proposed basis for modelling the dependence among different experts' densities is a recalibration function, based on the probability integral transforms of the expert densities. We show that this reduces to a copula function in a special case. We explore the properties of various approximations to the recalibration function both via Monte Carlo simulations and in an application density forecasting U.K. inflation using the Bank of England's "fan" chart. We find that the copula opinion pool can deliver more accurate densities than traditional linear and logarithmic opinion pools in many realistic situations when historical data on expert performance are available.

Subject Classifications: Expert Resolution; Density Forecast Combination; Opinion Pools; Dependence; Copula; Central Bank Fan Charts; Econometrics; Forecasting

*Address for correspondence: James.Mitchell@wbs.ac.uk. Thanks to seminar participants at Nottingham, Rome Tor Vergata, Rotterdam and Warwick (Economics) for helpful comments.

1 Introduction

Probability density forecasts are increasingly produced and used by decision makers in a range of environments including in business, economics, finance and meteorology. Density forecasts provide a full impression of the uncertainty associated with a forecast and in general facilitate *better* (i.e. lower cost) decisions.[1] Only when the Decision Maker has a quadratic loss function is their *optimal* (i.e. cost minimizing) decision unaffected by uncertainty (e.g. see Zellner (1986)). Density forecasts can be produced in many ways, reflecting the Decision Maker's objectives. A common approach is to employ some form of model, whether data or theory driven. But the density forecasts may also be elicited from subjective surveys or involve some judgement being applied to model-based forecasts, as with Delphi forecasts and committee-based forecasts. Or they may come from other sources. Our focus is on combining these density forecasts, taking them as given and assuming no feedback between Experts and the Decision Maker, irrespective of their informational source. They are simply, in the language of decision analysts, "Expert" density forecasts.[2]

There is a long tradition in operations research and management science of aggregating Experts' densities - of Expert resolution or consensus. Various methods of combining or aggregating these densities have been proposed, including axiomatic and modeling (or mathematical) approaches; see Winkler (1986), Genest & Zidek (1986) and Clemen & Winkler (1999) for reviews. This paper considers this model based approach. In its most common manifestation this involves use of the linear and logarithmic opinion pools. A more recent literature in econometrics has shown that (linear) pools across Expert densities can be effective in the face of temporal changes in Expert performance, often due to structural change (e.g., see Jore et al. (2010)), and when the Decision Maker is unsure which Expert is best and suspects that all Expert densities are likely incorrect (see Geweke & Amisano (2011)). This literature follows the spirit of Cooke's classical method (see Cooke (1991)) and like our proposed approach is designed for use in situations where the Decision Maker uses objective (historical or *seed variable*) data rather than subjective judgments to evaluate and aggregate the Experts' densities. The aim is for the Decision Maker to maximize the informational content of the combination given both the Experts' densities and data on their track-record in our case assumed to comprise the

[1] The terms risk and uncertainty are treated interchangeably here and, under a Bayesian (Savagian) interpretation, represented by a probability distribution.

[2] In the statistical literature these "Expert" densities are often called "component" densities; e.g. see Ranjan & Gneiting (2010).

Experts' historical density forecasts and the (*ex post*) realizations of the forecast variable of interest. In particular, assuming these data are available, combination weights can be tuned to reflect Experts' historical performance (e.g. see Hall & Mitchell (2007)). The linear opinion pool has shown promise in applications forecasting macroeconomic variables (e.g. Mitchell & Hall (2005), Hall & Mitchell (2007) and Garratt et al. (2011)), stock returns (e.g. Geweke & Amisano (2011)), meteorology (e.g. Raftery et al. (1995) and Ranjan & Gneiting (2010)) and for the Expert panels held at TU Delft (see Cooke & Goossens (2008)). Increasingly, in fact, large amounts of historical data on Experts' performance are available. These data could take the form of panel Expert elicitation surveys like those managed at TU Delft or the Survey of Professional Forecasters, held at the Federal Reserve Bank of Philadelphia, which contains quarterly macroeconomic Expert density forecasts from 1968. Or they could involve treating statistical models as Experts; with the Decision Maker fitting competing models to their data reflecting model uncertainty. In many time-series applications (e.g., in business, economics and finance) historical, perhaps simulated, forecast errors can play the role of the seed variables used to fit the opinion pool. The use of multiple (judgment-free) statistical models as Experts is common in many decision making contexts. While their combination is not necessarily a substitute for human Experts, it does summarize the information in the data about the forecast variable of interest.[3] Indeed, in the application below, the Bank of England consult a *suite* of statistical models to inform, albeit not exclusively determine, their fan chart for inflation (see Kapetanios et al. (2008)).

However, with a few notable exceptions which we discuss below and form the basis for our contribution, little explicit attention when combining these density forecasts has been paid to their possible dependence; this is the case for the widely used linear (and logarithmic) opinion pools. This apparent neglect is despite the shared information which many Experts condition on when forming their density forecasts; e.g. they often exploit overlapping information sets, use similar models to process common data and/or if the densities are formed subjectively read the same newspapers, etc. It also contrasts a larger literature, since Bates & Granger (1969), concerned with the combination of point forecasts accommodating their (linear) dependence.

We seek to remedy this omission and to do so draw on an established Bayesian paradigm, namely the Expert combination model of Morris (1974, 1977). Under this

[3]In many economic applications simple statistical forecasts are often found to perform at least on a par with both more complicated statistical or economic models and judgment based forecasts; e.g. see Clements & Hendry (1998).

approach, which might in practice involve use of frequentist estimation methods, the Decision Maker uses the Experts' densities as "data" to update, via Bayes' Theorem, their prior distribution about the future values of the variable(s) of interest. But to arrive at their posterior density, the Decision Maker is required to specify the joint density, or likelihood, derived from the Experts' marginal densities - the "data" as far as the Decision Maker (or meta Expert) is concerned. This requirement appears to have hindered application of the approach in practice: Morris' (1977, p. 687) words, written thirty-five years ago, remain true today: "One of the future challenges of Expert modeling is the construction of general models of Expert dependence". Instead we have seen specific applications, assuming normality as in Winkler (1981). Moreover, the focus in the literature, including recently in econometrics, has been on linear and logarithmic pools, where dependence is not accommodated, certainly explicitly.

In Section 2 we develop the Recalibrated and Copula Opinion Pools that accommodate, in a practical manner which is shown to be operational, any stochastic dependence between different Experts' densities. The basis for modeling dependence among different Experts' densities is a re-calibration function, based on the probability integral transforms of the Experts' densities - their "density forecasting errors" (cf. Mitchell & Wallis (2011)). We explain that this Recalibrated Opinion Pool reduces to the product of the Experts' densities and a copula function in a special case. Thereby, we explicitly relate Morris (1974, 1977) to Jouini & Clemen (1996) and develop the suggestion of Jouini & Clemen (1996) for the Decision Maker to use copula functions to link together the Experts' (marginal) densities to construct the multivariate density. By decoupling the model of stochastic dependence, captured by the copula, from the Experts' (marginal) densities we show that the Copula Opinion Pool can generate flexible posterior predictive densities, irrespective of the distributional form of the Experts' densities. By exploiting the probability integral transforms, rather than just the point forecasting errors as in Jouini & Clemen (1996), the Copula Opinion Pool (COP), via the choice of the copula function, accommodates not only linear but asymmetric dependence too. Pearson correlation offers a sufficient measure of dependence only under joint normality. In turn, we show explicitly how the COP generalizes the approach to handling dependence suggested by Winkler (1981), which is based on the multivariate normal distribution and works off the point forecasting errors only; with the point forecast usually defined as the conditional mean of the predictive density.

More generally, drawing on recent developments in econometrics, in Section 3 we consider how to operationalize the ROP and in particular how to estimate the COP. Since in many applications the Decision Maker has a time-series of (historical) forecasts from the Experts, which can be evaluated against the subsequent outturns, we consider optimal estimation of the COP using these time-series data. Optimality is defined generally, free from specific assumptions about the nature of the user's loss function, with respect to the average logarithmic score, generalizing Hall & Mitchell (2007) and Geweke & Amisano (2011) who consider linear combinations only of the Experts. The optimal combination density forecast is the one that maximizes the logarithmic predictive score. Thereby past performance of the pool, over a training period, is used to determine the nature of the COP.

Then in Section 4 we undertake Monte Carlo simulations to show that dependence can have serious effects on the nature of the combined density. We find the COP offers gains relative to the linear and logarithmic opinion pools. Section 5 then reinforces the superiority of the COP in an application forecasting U.K. inflation using the Bank of England's fan chart. Section 6 concludes.

2 Dependence among the Experts' densities

Typically dependence between point forecasts and point forecast errors is captured by (Pearson) correlation. However, correlation offers a sufficient summary measure of association only when the point forecasts are jointly normally or, more generally, elliptically distributed. Moreover, while correlation captures linear dependence it cannot measure nonlinear or asymmetric dependence.[4]

Therefore, a more general approach is required for Expert (marginal) forecast densities which may not be Gaussian and even when Gaussian may not follow a multivariate normal distribution. Moreover, we may well expect some type of asymmetric dependence; e.g., there may be differing degrees of correlation between forecasts during upturns than downturns or in times of pronounced uncertainty. Indeed, in the application below forecasting U.K. inflation, we find that the two Experts' forecasts are more dependent when the outturn falls in the middle of the distribution - when, in a sense, it is business-as-usual - than in the tails, when an extreme event occurs. Pearson (linear) correlation is unable to detect asymmetric dependence like this.

We might also hope to obtain "better" combined density forecasts, or opinion pools, if we account for (any) dependence between N Experts. Certainly, it is well known that the

[4]See Embrechts et al. (1999) for a description of the dangers associated with the use of correlation as a measure of dependence.

"optimal" (i.e., mean squared error (MSE) minimizing) in-sample combination of point forecasts involves examination of the dependence between the competing forecasts; see Bates & Granger (1969). Only when the forecasts are uncorrelated do the optimal weights equal the inverse of the relative root MSEs.

Despite this expectation that accommodating Expert dependence will be beneficial, the two most popular means of combining density forecasts do not account, at least explicitly, for dependence:

1. The Linear Opinion Pool (LOP) takes a weighted linear combination of the Experts' probabilities
$$p(y)^{LOP} = \sum_{i=1}^{N} w_i g_i(y) \tag{1}$$
where $g_i(y)$ are the (conditional) density forecasts of Expert i, where $\sum_{i=1}^{N} w_i = 1$.

2. The Logarithmic Opinion Pool (logOP) is
$$p(y)^{\log OP} = k \prod_{i=1}^{N} g_i(y)^{w_i} \tag{2}$$
where k is a normalizing constant. The logOP, like any geometric combination including those proposed below, will assign a zero probability to any y value if for at least one Expert, i, $g_i(y) = 0$. When Expert densities have support over the extended real line $(-\infty, \infty)$, as they would for Gaussian densities for example, this will not happen. Thus an important issue for the Decision Maker is selecting which N Experts to pool.

2.1 The Bayesian approach to the aggregation of density forecasts

Our proposed method for handling Expert dependence draws on an established methodological framework for the combination of competing forecasts, namely the Bayesian approach to the aggregation of density forecasts. Following Morris (1974, 1977), Bayes' Theorem is used to update the Decision Maker's prior distribution of the variable y, $d(y)$, in the light of "data" from the $i = 1, ..., N$ Experts that takes the form of the joint density, or likelihood, derived from their N densities. (While an Expert may view their density forecast as reflecting what they know, it is the Decision Maker's interpretation of the Experts' densities as information that enables Bayes' Theorem to be invoked.) This delivers the posterior density of y conditional on the Decision Maker's prior and the Experts' (conditional) densities $g_i(y)$:

$$p(y|g_1, ..., g_N) = k.f(g_1, ..., g_N|y)d(y) \tag{3}$$

where k is a normalization constant and $f(g_1, ..., g_N|y)$ is the likelihood function of the Experts' predictions. Following Jouini & Clemen (1996) we assume that everything the Decision Maker knows about y is captured by the Experts' densities and adopt a non-informative prior. If the Decision Maker did have more information we could simply capture this with another density, the $N + 1$-th Expert's density. We therefore ignore $d(y)$ below. We also assume no feedback or interaction between the Experts and the Decision Maker, at least after the Expert has supplied their forecast.

The difficulty faced by the Decision Maker when implementing this approach is deciding upon the form of the likelihood function or joint density. The likelihood must capture the bias and precision of the Experts' densities as well as their dependence, a point also made by Clemen (1986). To-date this has precluded widespread application of this method. As discussed in the Introduction we focus on situations where the Decision Maker uses objective (historical or *seed variable*) data rather than subjective judgments to specify $f(g_1, ..., g_N|y)$. The aim is for the Decision Maker to maximize the informational content of the combination given both the Experts' densities and data on their track-record. In our case the latter is assumed to comprise the Experts' historical density forecasts as well as the (*ex post*) realizations of the forecast variable of interest. Thereby our approach to Expert combination is deliberately pragmatic with objective historical data on Expert performance used to specify the joint density from possibly subjective Expert densities.

2.2 Winkler (1981): combining densities looking at their first two moments only

One popular way to implement the Bayesian approach, (3), is to follow Winkler (1981) and assume the multivariate distribution is normal and characterize dependence based on analysis of the point forecasting errors only. (Lindley (1983) derives an analogous result but from a different starting point.) This delivers a tractable analytical expression for

$f(g_1, ..., g_N|y)$.

Let $m_i = \int_{-\infty}^{\infty} y g_i(y) dy$ and $v_i = \int_{-\infty}^{\infty} (y - m_i)^2 g_i(y) dy$, respectively, denote Expert i's (conditional) mean and variance forecast of y, with $\mathbf{m} = (m_1, ..., m_N)'$ and $\mathbf{v} = (v_1, ..., v_N)'$.[5] The forecasting error for Expert i is $s_i = m_i - y$.

Assume the N-stacked vector $\mathbf{s} = (s_1, ..., s_N)'$ is mean zero with known covariance matrix $\mathbf{\Sigma}$, where the diagonal of $\mathbf{\Sigma}$ comprises $\mathbf{v}$, thereby leaving $N(N-1)/2$ remaining elements of $\mathbf{\Sigma}$ to be specified. The forecasts m_i can be recalibrated prior to combination if biased. In addition, and we consider this below, the diagonal of $\mathbf{\Sigma}$ could be estimated as $E(\mathbf{ss}')$. This need not deliver the same estimates for $\mathbf{v}$ if they are not conditional expectations.

Assuming $f(y|g_1, ..., g_N) \propto f(s_1, ..., s_N|y)$ and $\mathbf{s} \sim N(\mathbf{0}, \mathbf{\Sigma})$, such that f is a N-variate normal density, Winkler (1981) shows that the posterior density for y, the combined density (3), conditional on the mean forecasts $\mathbf{m}$ (and possibly but not necessarily $\mathbf{v}$ too), reduces to:

$$p(y|\mathbf{m}) \propto \phi\left[(y - m^*)/\sigma_m^*\right], \tag{4}$$

where ϕ is the standard normal density function and

$$m^* = \mathbf{e}'\mathbf{\Sigma}^{-1}\mathbf{m}/\mathbf{e}'\mathbf{\Sigma}^{-1}\mathbf{e}. \tag{5}$$

$$\sigma_m^{*2} = 1/\mathbf{e}'\mathbf{\Sigma}^{-1}\mathbf{e}. \tag{6}$$

m^* is the combined point forecast, which we discuss further below. It can be seen from a frequentist perspective, given the assumed non-informative prior, as the maximum likelihood (ML) estimator.[6] In this sense m^* is optimal. It is equivalent to the minimum variance solution of Bates & Granger (1969). Below we show how m^* is not just MSE optimal but the circumstances in which it is the mean of the (logarithmic score) optimal combined density forecast. Winkler (1981) also shows that when $\mathbf{\Sigma}$ is unknown, and determined from a prior (the inverted Wishart distribution) or estimated from a sample covariance matrix, $h(y|\mathbf{m})$ is a t-density, with the same posterior mean m^* as in (5) and

[5] Given non-quadratic loss functions it is well known that the "optimal" central estimate may not equal the mean; e.g. see Zellner (1986).

[6] The likelihood $h(\mathbf{m}|y)$ implies $\mathbf{m}$ are N random variables sampled from a normal distribution with common mean y and variance-covariance matrix $\mathbf{\Sigma}$. Treating $\mathbf{\Sigma}$ as fixed and known, differentiation of the likelihood with respect to y reveals that the ML estimator m^*: $\sqrt{T}(m^* - y) \xrightarrow{d} N(0, (\mathbf{e}'\mathbf{\Sigma}^{-1}\mathbf{e})^{-1})$. See also Halperin (1961).

posterior variance a scale multiple of σ_m^{*2}. As the degrees of freedom increase this scalar factor trends to unity, and the posterior density tends to normality - as in (4).

Moreover, this approach characterizes dependence based on analysis of the point forecasting errors only, when we might wish to model the dependence between all fractiles of the densities.[7] It does not exploit any available information on other characteristics of the density forecasts. In contrast, we propose methods that combine known density forecasts not just in terms of the accuracy of the point forecasts but the calibration of the 'whole' density as indicated by their probability integral transforms. In so doing we show that the copula approach is easy to apply when one does more than look at point forecast errors as in Jouini & Clemen (1996).

2.3 The Recalibrated Opinion Pool (ROP)

Without loss of generality consider $N = 2$ Experts. The posterior density of y conditional on the Experts' densities $g_1(y)$ and $g_2(y)$ is given, via Bayes' Theorem, as

$$p(y|g_1, g_2) = k.f(g_1, g_2|y), \tag{7}$$

where k is a normalization constant and $f(g_1, g_2|y)$ is the likelihood function of the Experts' predictions. This represents the Decision Maker's model of the Experts and is a joint probability assessment conditional on the true (but in practice often unknown as we explain below) realization y.

Importantly, following Morris (1977) eqn. 25, $f(g_1, g_2|y)$ can be re-written as the product of the marginal densities and a "joint re-calibration function" $c(z_1, z_2)$, where $z_1 = G_1(y)$ and $z_2 = G_2(y)$ are the probability integral transforms (pits), with $\frac{dG_1(y)}{dy} =$

[7] Jouini & Clemen (1996) also work off the point forecasts only, even though the copula approach they describe is, as we explain below, more general.

$g_1(y)$ and $\frac{dG_2(y)}{dy} = g_2(y)$. This yields the ROP:[8]

$$p(y|g_1, g_2) = k.c(z_1, z_2).g_1(y).g_2(y). \tag{8}$$

In practice (although we continue to drop temporal subscripts when notationally convenient below) it is important to distinguish the temporal flow of information in this Expert aggregation problem. Define a time-series sequence $t = 1, ..., T$, with $g_{iT}(y_{T+1}) = g_{iT+1}$ denoting the i-th Expert's one-period ahead forecast of y_{T+1} formed using information dated up to and including period T. The h-period ahead ($h > 1$) forecast can be defined analogously. The ROP for y_{T+1}, conditional on the Experts' density forecasts for y_{T+1}, formed at time T, and both realizations and forecasts up to period T, is then defined as

$$p(y_{T+1}|g_{1T+1}, g_{2T+1}, \{g_{1t}, g_{2t}, y_t\}_{t=1}^{T}) = k. \{c(z_{1t}, z_{2t})\}_{t=1}^{T} .g_{1T}(y_{T+1}).g_{2T}(y_{T+1}), \tag{9}$$

indicating that the joint re-calibration function is only defined using pits data available to the Decision Maker, which given that y_{T+1} is not published until at least (depending on publication lags) period $T + 1$ does not include z_{iT+1}. The Decision Maker only learns z_{iT+1}, conveniently a one-dimensional object unlike g_{iT+1}, on receipt of y_{T+1}.

2.4 Properties of the ROP: the Copula Opinion Pool as a special case

Specification of $f(g_1, g_2|y)$ is key to combining dependent Expert densities. As (8) shows it contains information beyond that contained in $g_1(y)$ and $g_2(y)$. Specifically, the joint re-calibration function $c(z_1, z_2)$ in (8) reflects both the "probabilistic calibration" of each

[8]Morris (1977) sets out the required scale and shift invariance assumptions. Assuming the mapping from y to z_1 and z_2, conditional on the Experts' densities, is known and one-to-one a change of variables such that

$$\begin{aligned}
c(z_1, z_2) &= f(G_1^{-1}(z_1), G_2^{-1}(z_2)) \begin{vmatrix} \frac{dG_1^{-1}(z_1)}{dz_1} & \frac{dG_1^{-1}(z_1)}{dz_2} \\ \frac{dG_2^{-1}(z_2)}{dz_1} & \frac{dG_2^{-1}(z_2)}{dz_2} \end{vmatrix} \\
c(z_1, z_2) &= f(G_1^{-1}(z_1), G_2^{-1}(z_2)) \begin{vmatrix} \frac{1}{g_1(y)} & 0 \\ 0 & \frac{1}{g_2(y)} \end{vmatrix} \\
f(g_1, g_2) &= g_1(y).g_2(y)c(z_1, z_2).
\end{aligned}$$

helps motivate the ROP. In reality, as we discuss in (9) below, the Decision Maker does not learn z_i until they know y. In any case our focus is on understanding use of the ROP - and in turn the COP - in practice rather than formal derivations from first principles.

Expert's density and Expert dependence.[9]

Probabilistic calibration of the forecast distribution $G_{it}(y)$ relative to the true (but in general unknown) true distribution $M_t(y)$ is defined, for the sequence $t = 1, ..., T$, by Gneiting et al. (2007) as

$$\frac{1}{T}\sum_{t=1}^{T} M_t G_{it}^{-1}(z_i) \to z_i \text{ for all } z_i \in (0,1) \tag{10}$$

$(i = 1, ..., N)$ and indicates marginal density calibration failure when the pits $z_{it} = G_{it}(y)$ deviate from uniformity (see Theorem 2 in Gneiting et al. (2007)). It does not, however, capture calibration failure resulting in serial dependence of the pits as explained in Mitchell & Wallis (2011). Two different Experts, with different density forecasts, can both satisfy (10) and deliver uniform pits, even in large samples, if they correctly condition on their differing and incomplete information sets; in this case misscalibration is picked up via temporal dependence of the z's.

We delineate special cases of the ROP, (8), by rewriting the joint density $c(z_1, z_2)$ as the product of its marginals and a copula function $c^*(.)$:

$$c(z_1, z_2) = f_1(z_1) f_2(z_2) c^*(F_1(z_1), F_2(z_2)) \tag{11}$$

where $\frac{dF_i(z_i)}{dz_i} = f_i(z)$ and $c^*(.)$ can capture, as we review in Section 3.1 below, general forms of Expert dependence. This decomposition, possible for any multivariate density, follows from Sklar's theorem (Sklar (1959)).

When the marginal densities $g_1(y)$ and $g_2(y)$ are probabilistically well-calibrated and their pits are uniform $f_1(z_1) = 1$, $F_1(z_1) = z_1$ and $f_2(z_2) = 1$, $F_2(z_2) = z_2$. In this case

$$c(z_1, z_2) = c^*(z_1, z_2) \tag{12}$$

and the Copula Opinion Pool (COP) is given as

$$p(y|g_1, g_2) = k.c^*(z_1, z_2).g_1(y).g_2(y). \tag{13}$$

[9] Recalibration functions have also been employed to improve combined density forecasts in cases where Expert dependence is ignored. Outside of the Bayesian framework, Ranjan & Gneiting (2010), for example, use the linear opinion pool to combine the densities and then recalibrate the pool based on their pits, z. In our framework, a variant of this would involve combining the densities using the logarithmic opinion pool with unit weights, $g_1(y).g_2(y)$, and recalibrating not using $c^*(z_1, z_2)$ which accommodates dependence as well as calibration, but $c^*(z)$ which reflects calibration only.

The COP is determined by $c^*(z_1, z_2)$, which is a copula function. This amounts to the case considered by Jouini & Clemen (1996), who do not discuss calibration, or lack of. Jouini & Clemen (1996) also focused on specific ways to estimate $c^*(z_1, z_2)$, based on the point forecasting errors, and did not exploit the "density forecasting errors", z_i.

But when z_1 and z_2 are not both uniform they are re-calibrated:

$$c(z_1, z_2) = f_1(z_1)f_2(z_2)c^*(1 - F_1(z_1), 1 - F_2(z_2)) \tag{14}$$

and essentially we need to model the multivariate density of the uniform pits $c(z_1, z_2)$. But this can be seen as akin to re-calibrating the marginal densities, using the re-calibration functions f_1 and f_2, and then modelling their dependence via the copula.[10]

Only under Expert independence does $c(z_1, z_2) = 1$. Then the COP reduces to the product of individual densities as in a logarithmic pool, (2), but with unit weights on each individual density.

2.5 Familiar special cases of the COP

2.5.1 Gaussian Expert densities and a Gaussian copula ⇒ Winkler (1981)

Let $g_1(y)$ and $g_2(y)$ be two Expert Gaussian densities with means m_i and standard deviations σ_i $(i = 1, 2)$. Then the COP is given as

$$p(y|g_1, g_2) = k.c(z_1, z_2). \prod_{i=1}^{2} \frac{1}{\sigma_i}\phi(u_i) \tag{15}$$

where $u_i = (y - m_i)/\sigma_i$ and ϕ is the p.d.f. of the standard normal distribution.

When $c(z_1, z_2) = c(G_1(y), G_2(y)) = c(\Phi(u_1), \Phi(u_2))$ is a Gaussian copula p.d.f. it takes the form

$$\frac{1}{|R|^{1/2}} \exp\left\{-\frac{1}{2}u'(R^{-1} - I)u\right\} \tag{16}$$

where $u = (u_1, ..., u_N)'$, $u_i = \Phi^{-1}(G_i(y)) = (y - m_i)/\sigma_i$ and $R = \{r\}$ is the correlation matrix of the standardized point forecasting errors, the u_i's, such that $\mathbf{\Sigma} = \mathbf{DRD}$ where $\mathbf{D} = diag\{\sigma_i\}$.

[10] In this case, it might be helpful to parameterize this process. One possibility is to use the beta density (since this keeps the outturns between 0 and 1). This is the density analogue of the beta-transformed combination in Ranjan & Gneiting (2010).

Then the normal COP is

$$p(y|g_1, g_2)^{NorCOP} = \frac{1}{|R|^{1/2}} \exp\left\{-\frac{1}{2}u'(R^{-1} - I)u\right\} \prod_{i=1}^{N} \frac{1}{\sigma_i}\phi(u_i) \tag{17}$$

where the Experts' (marginal) densities are fixed.

But since

$$\frac{1}{|R|^{1/2}} \exp\left\{-\frac{1}{2}u'(R^{-1} - I)u\right\} \prod_{i=1}^{N} \frac{1}{\sigma_i}\phi(u_i) = \frac{\exp\left\{-\frac{1}{2}u'R^{-1}u\right\}}{(2\pi)^{N/2} \prod_{i=1}^{N} \sigma_i \, |R|^{1/2}} \tag{18}$$

it follows that $p(y|g_1, g_2)^{NorCOP}$ is equivalent to Winkler seen in (4) as $\frac{\exp\left\{-\frac{1}{2}u'R^{-1}u\right\}}{(2\pi)^{N/2} \prod_{i=1}^{N} \sigma_i |R|^{1/2}} = \frac{\exp\left\{-\frac{1}{2}(y-\mathbf{m})'\mathbf{\Sigma}^{-1}(y-\mathbf{m})\right\}}{(2\pi)^{N/2}|\mathbf{\Sigma}|^{1/2}}$ is nothing more than the multivariate normal density, on which Winkler relies. So R is the correlation matrix of the point forecasting errors, but in this special case, given the assumed Gaussian copula, this fully captures the dependence between the density forecasting errors too.[11] This means we know $p(y|g_1, g_2)^{NorCOP}$ is Gaussian with mean and variance as given in (5) and (6).

When R is diagonal $\frac{1}{|R|^{1/2}} \exp\left\{-\frac{1}{2}u'(R^{-1} - I)u\right\} = 1$ and $p(y|g_1, g_2)^{NorCOP}$ reduces to a logarithmic opinion pool.

An important distinction is that, focusing on $N = 2$, in Winkler there are three parameters to estimate in $\mathbf{\Sigma}$ but only one in norCOP - the correlation coefficient. That is, Winkler estimates all of $\mathbf{\Sigma}$; i.e. it essentially re-calibrates the variances of the marginal densities so that they equal the variance of the point forecast errors. This need not equal the variance of the density forecast, when the density forecast is not constructed as the efficient projection so that the pits are uniform (and the point forecast errors are unbiased). In the norCOP these variances are fixed, and only the correlation coefficient is estimated.

[11] Clemen & Reilly (1999) provide a discussion of the use of the normal copula when modelling multivariate processes rather than multiple Expert forecasts of a scalar, y, as here.

3 Estimation of the ROP and the COP

Our focus, as explained in the Introduction, is on estimating the parameter(s) of the opinion pools objectively using historical data. Alternative approaches include subjective assessment of the parameters; e.g. see Clemen & Reilly (1999).

The COP is more convenient to operationalize and estimate than the ROP. The ROP, (8), with the Experts' densities taken as given, requires the recalibration function $c(z_1, z_2)$ to be be specified. Specification of this appears as demanding as of $f(g_1, g_2|y)$ itself. Any bivariate (more generally multivariate) distribution on the unit circle is permissible; and it is unclear when and how the Decision Maker might have a preference for a specific (known) distribution. In the simulations below, we therefore consider nonparametric estimators for $c(z_1, z_2)$. This, of course, offers a practical means of estimating the ROP in large samples only.

But with the marginal densities $g_1(y)$ and $g_2(y)$ considered probabilistically well-calibrated, perhaps after post-processing, the COP 'only' requires specification of the copula function $c^*(z_1, z_2)$. Again nonparametric estimators might be considered, but given their possible advantages in small samples and to contrast the ROP, we consider a parametric approach. This requires a copula function to be chosen.

3.1 Choice of the copula function

The practical obstacle to easy implementation of the COP is choosing which copula to use. As Joe (1997) and Nelsen (1999) review, there are many to choose from. We confine attention here to the normal, t, Joe-Clayton and Frank copulae as they are sufficient to illustrate the flexibility of the COP.

Different copula functions allow specifically for different types of dependence; they allow for the fact that association may be stronger in one part of the distribution than the other. The issue is to determine the 'right' copula. Our use of scoring rules to evaluate performance, see section 3.2 below, suggests their use when selecting which copula function to use in the COP. As section 4 then goes on to explain, Kullback-Leibler tests for equal predictive performance, constructed between COPs with different copula functions, provide a means of constructing selection tests. Our focus in this paper, in illustrating use of the COP, is on the use of static copula functions; dynamic functions could be of interest in future time-series applications.

To illustrate use of the COP in the simulations and application that follow we entertain four copulae. The first is well known and is the normal copula, seen in (16), determined by a single parameter r, where $r \in (-1, 1)$.

The second is the t copula density, determined by two parameters r and $\widetilde{v}$, $\widetilde{v} > 2$. Similarly to the univariate case, the t-copula generalizes the normal copula by allowing for joint fat tails; this means it allows for a higher probability of extreme outcomes for both marginal densities. The t-copula allows the two densities to be related in the extreme tails even when $r = 0$. For the normal copula there is zero tail dependence as long as $r < 1$; see Embrechts et al. (1999). Therefore, as Chen et al. (2004) show, the differences between the normal and t-copulae can be significant - they can imply quite different dependence characteristics.

Third is the Joe-Clayton copula, determined by two parameters τ^U and τ^L, $\tau^U, \tau^U \in (0, 1)$, and given as $c_{JC}(z_1, z_2) =$

$$1 - \left(1 - \left\{[1-(1-z_1)^\kappa]^{-\gamma} + [1-(1-z_2)^\kappa]^{-\gamma} - 1\right\}^{-1/\gamma}\right)^{1/\kappa} \tag{19}$$

where $\kappa = 1/\log_2(2 - \tau^U)$, $\gamma = -1/\log_2(\tau^L)$. τ^U and τ^L allow for tail dependence; and thereby the Joe-Clayton copula accommodates extreme events, such as both Experts making large density forecasting errors in either the same or opposite directions. The normal copula, when $r < 1$, has $\tau^U = \tau^L = 0$, implying that both Experts are independent in the tails. We follow Patton (2006) and to impose symmetry of the copula when $\tau^U = \tau^L$ use the symmetrized Joe-Clayton copula $c_{SJC}(z_1, z_2) =$

$$0.5c_{JC}(z_1, z_2) + 0.5c_{JC}(1 - z_1, 1 - z_2) + z_1 + z_2 - 1 \tag{20}$$

Fourth is the Frank copula, determined by $\theta \in (-\infty, \infty)$, and given as $c_{Frank}(z_1, z_2) =$

$$-\frac{1}{\theta} \ln \left(1 + \frac{(\exp\{-\theta z_1\} - 1)(\exp\{-\theta z_2\} - 1)}{\exp\{-\theta\} - 1}\right) \tag{21}$$

and implies asymptotic tail independence.

3.2 Optimal estimation of the COP

Given a time-series ($t = 1, ..., T$), the past performance of the opinion pool - over some historical training period - can be used to estimate the parameters in the COP. This use of historical data is similar both to how point forecast are combined (see Bates & Granger (1969)) and to how the weights are chosen in *optimal* linear opinion pools (see Hall &

Mitchell (2007) and Geweke & Amisano (2011)).

Optimality is defined generally, free from specific assumptions about the nature of the user's loss function, with respect to the average logarithmic score, generalizing Hall & Mitchell (2007) and Geweke & Amisano (2011) who consider linear combinations of the Experts only.[12]

The optimal COP is the one that maximizes the logarithmic predictive score. As Hall & Mitchell (2007) discuss, by maximizing the logarithmic score of the COP its Kullback-Leibler Information Criterion (KLIC) *distance* relative to the true but unknown density is being minimized. As in Geweke & Amisano (2011), no assumption is made in estimation that one of the Experts is correct.

Specifically, the $KLIC$ distance between the true density $m_t(y_t)$ and the copula opinion pool $p_t(y_t)$ $(t = 1, ..., T)$ is defined as:

$$KLIC_t = \int m_t(y_t) \ln \left\{ \frac{m_t(y_t)}{p_t(y_t)} \right\} dy_t \text{ or} \tag{22}$$

$$KLIC_t = \mathsf{E}\left[\ln m_t(y_t) - \ln p_t(y_t)\right]. \tag{23}$$

The smaller this distance the closer the density forecast to the true density. $KLIC_t = 0$ if and only if $m_t(y_t) = p_t(y_t)$.

Under some regularity conditions $\mathsf{E}\left[\ln m_t(y_t) - \ln p_t(y_t)\right]$ can be consistently estimated by $\overline{KLIC}$, the average of the sample information on $m_t(y_t)$ and $p_t(y_t)$ $(t = 1, ..., T)$:

$$\overline{KLIC} = \frac{1}{T} \sum\nolimits_{t=1}^{T} \left[\ln m_t(y_t) - \ln p_t(y_t)\right]. \tag{24}$$

Definition 1 *The optimal COP is $p^*(y|g_1, g_2)$, where the optimal parameter vector $\boldsymbol{\rho}_T^*$ minimizes this KLIC distance. This minimization is achieved as follows:*

$$\boldsymbol{\rho}_T^* = \arg\max_{\boldsymbol{\rho}} h_T(\boldsymbol{\rho})$$

where $h_T(\boldsymbol{\rho}) = \frac{1}{T} \sum\nolimits_{t=1}^{T} \ln p_t(y_t|g_{1t}, g_{2t})$ is the average logarithmic score of the COP over the training sample $t = 1, ..., T$.

[12] Gneiting & Raftery (2007) discuss a general class of proper scoring rules to evaluate density forecast accuracy, whereby a numerical score is assigned based on the predictive density at time j and the value of y that subsequently materializes, here assumed without loss of generality to be at time $j+1$. A common choice for the loss function L_T, within the proper class (cf. Gneiting & Raftery (2007)), is the logarithmic scoring rule. More specific loss functions might be appropriate in some applications. In this case these rather than the logarithmic score might be minimized via the following.

Assuming concavity for $h(\boldsymbol{\rho})$, $\boldsymbol{\rho}_T^* = \arg\max_{\boldsymbol{\rho}} h_T(\boldsymbol{\rho})$ converges almost surely to $\boldsymbol{\rho}^* = \arg\max_{\boldsymbol{\rho}} h(\boldsymbol{\rho})$.

4 Monte Carlo Simulations

To explore the properties of the ROP and COP, and compare them to the LOP and logOP alternatives which do not accommodate Expert dependence, we carry out a set of Monte Carlo simulations. (These simulations, which indicate how dependence has a serious effect on the nature of the combined density, are consistent with simple experiments (not reported) we carried out involving use of the COP to combine two different Gaussian Expert densities using the four different copula functions considered above. These simple experiments confirmed that the COP can generate more flexible densities, with skewness and kurtosis, when we move beyond the normal copula and vary the parameter(s) in the copula.)

In each case the KLIC is used to judge density forecasting performance of the respective pool, $p_{jt}(y_t)$, relative to the true or ideal (i.e. the correct) conditional density, $m_t(y_t)$:

$$\mathsf{KLIC}_{jt} = E\left\{\ln m_t(y_t) - \ln p_{jt}(y_t)\right\} = E\left\{d_{jt}(y_t)\right\}. \tag{25}$$

KLIC_{jt} is the expected difference in their log scores, with $d_{jt}(y_t)$ the "density forecasting error" (Mitchell & Wallis (2011)), which can be used to construct Giacomini & White (2006) type tests for equal predictive accuracy $m_t(y_t) = p_{jt}(y_t)$.

To ensure relevance in a time-series forecasting context we forecast an autoregressive process using two (misspecified) statistical Experts. This involves extending the simulation experiments in Smith & Wallis (2009) and Mitchell & Wallis (2011) who, not modelling Expert dependence, focus on the LOP and logOP with equal combination weights. Here we seek to isolate how Expert dependence affects the relative performance of the different Opinion Pools.

4.1 Forecasting an autoregressive process

Consider the second-order autoregressive data-generating-process

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t,\ \varepsilon_t \sim N(0, \sigma_\varepsilon^2). \tag{26}$$

The true or 'ideal' forecast distribution of Y_t given an information set Ω_t comprising observations y_{t-1} and y_{t-2}, the model and its parameter values is

$$m_t(y_t) = N(\phi_1 y_{t-1} + \phi_2 y_{t-2}, \sigma_\varepsilon^2). \tag{27}$$

The following two (misspecified) Expert densities are then combined via the various opinion pools:

1. Expert 1 (AR1) is a random walk forecaster:

$$AR1_t = N(y_{t-1}, 2(1-\rho_1)\sigma_y^2), \tag{28}$$

where $\sigma_\varepsilon^2 = (1 - \phi_1\rho_1 - \phi_2\rho_2)\sigma_y^2$ and ρ_i, are autocorrelation coefficients:

$$\rho_1 = \phi_1/(1-\phi_2),\ \rho_2 = \phi_1\rho_1 + \phi_2. \tag{29}$$

2. Expert 2 (AR2) uses a first order AR with the same variance:

$$AR2_t = N((2\rho_1 - 1)y_{t-1}, 2(1-\rho_1)\sigma_y^2). \tag{30}$$

The contemporaneous correlation between the Expert's forecast errors is ρ_1. Expert dependence can therefore be controlled in the simulations below by varying ϕ_1 and ϕ_2 (subject to stationarity restrictions). Thereby, we establish how Expert dependence affects the performance, as measured by the KLIC, of the different opinion pools. If focus were on MSE loss - and the points forecasts only - since the variances of both Expert densities are identical, the optimal combination involves equal weights, irrespective of ρ_1. The simulations below therefore indicate if and how Expert dependence does matter when focus is on the entire density, not simply the conditional mean.

We assume that these Experts use least-squares regression of y_t on its lagged value to estimate the parameters in their densities, but we neglect parameter estimation error and use the corresponding 'true' values. Therefore, while misspecified, having ignored y_{t-2}, each Expert is probabilistically well-calibrated. While we take these Expert densities as given, the parameter(s) in the Opinion Pool are estimated at each Monte Carlo replication. While it might be hoped that some human Experts would learn from their (forecasting) mistakes, both of these model-based Experts persist in using the incorrect statistical model. While one could entertain statistical models either with learning or that are robust to misspecification, our simulation design is designed to draw out the effects of Expert dependence on the different opinion pools.

4.2 Simulation results

To distinguish time dependence from Expert dependence we vary ϕ_1 and ϕ_2, and in turn ρ_1 and ρ_2, to generate six different stationary processes used to estimate and fit the two Expert densities and the various opinion pools. We focus on $T = 150$, typical of many macroeconomic samples (and recall the parameter(s) in the opinion pools are estimated at each replication).[13] We carry out 1000 Monte Carlo replications for each experiment.

Table 1 shows the KLIC rejection rates at the nominal 5% level as time and Expert dependence change for each Expert and the various opinion pools. These include the optimized LOP (denoted LOPop) as well as the equal weighted LOP and logOPs, simply denoted LOP and logOP in Table 1.

We draw two main conclusions from Table 1. First, looking at the final three columns, with temporal dependence ρ_2 fixed at 0.4, we see that as ρ_1 is increased the rejection rates decrease. This improvement in performance for the pools as Expert dependence increases is most pronounced for the COPs, especially the Symmetrized Joe-Clayton copula which has the lowest rejection rates. The traditional linear and logarithmic pools, including when the combination weights in the LOP are optimized, do not match the performance of the COPs, with at $\rho_1 = 0.67$ the COPs offering gains of nearly 50%. Accommodating Expert dependence via the COP delivers improved density forecast accuracy.

Secondly, looking at the first three columns of Table 1, when Expert dependence is fixed at $\rho_1 = 0.33$ but temporal dependence as measured by ρ_2 is increased instead, we see rejection rates increase across the board. This reflects the fact that as ρ_2 rises the component Expert densities themselves become poorer. This deterioration in quality of the Expert densities becomes so pronounced that no opinion pool can deliver competitive density forecasts. Rubbish in, Rubbish Out. Hence we see rejection rates of 100% when $\rho_2 = 0.8$. But for lower values of ρ_2 we again see benefits to combination, with again

[13] We also carried out experiments for other T, finding minor deteriorations in performance for the COP as T declined below $T = 50$ (detailed results available upon request). The ROP results below are a cheat and use 10,000 (rather than $T = 150$) pits from both Experts: we then estimate the multivariate density of the pits $c(z_1, z_2)$ nonparametrically via the bivariate histogram with 100 bins in $[0, 1]^2$. We did, in an attempt to facilitate a fair comparison with the other pools, experiment when T is smaller with fitting a (normal) kernel but the ROP performed very poorly. We conclude that implementation of the ROP is practical only in very large samples. The ROP results below are therefore presented only for illustrative purposes.

Table 1: Simulation Results: KLIC rejection rates at nominal 5% level as time and Expert dependence change

	ρ_1=0.33 ρ_2=0.2	ρ_1=0.33 ρ_2=0.4	ρ_1=0.33 ρ_2=0.8	ρ_1=0.02 ρ_2=0.4	ρ_1=0.33 ρ_2=0.4	ρ_1=0.67 ρ_2=0.4
AR1	100	100	100	100	100	96
AR2	100	100	100	100	100	67
ROP	6	47	100	67	47	6
norCOP	9	51	100	60	51	12
sjcCOP	4	40	100	74	40	3
frankCOP	5	42	100	60	42	7
tCOP	14	55	100	62	55	8
LOPop	96	98	100	100	98	49
LOP	97	99	100	100	99	58
logOP	14	48	100	71	48	50

Notes: AR1 and AR2 are the two (marginal) Expert forecast densities; ROP is the recalibrated opinion pool; norCOP, sjcCOP, tCOP and frankCOP are the copula opinion pools with normal, Symmetrized Joe-Clayton, t and Frank copula functions; LOPop is the linear opinion pool with optimized weights; LOP is the equal-weighted linear opinion pool; logOP is the logarithmic opinion pool.

the COPs outperforming traditional linear and logarithmic pools. The linear pool in particular performs poorly, with rejection rates close to 100% even when $\rho_2 = 0.2$.

5 An application: the Bank of England's "fan" chart for inflation

Economic forecasts play a central role in helping the Monetary Policy Committee at the Bank of England assess the key economic risks when they set monetary policy. We consider the quarterly sequence of inflation forecasts published by the Bank of England in their *Inflation Report* in February, May, August and November, which we correspond to quarters q1, q2, q3 and q4, respectively. These forecasts are not mechanically produced by a model or combination of models but also reflect the Bank's judgment.

The Bank of England has published density forecasts for inflation, at least up to eight quarters ahead, from 1993q1. Up until 1995q4 these took the form of charts showing the central projection, together with an estimate of uncertainty based on the historical mean absolute error. At this stage the Bank of England did not quantify a skew so that modal, median and mean projections are equal; the density forecast is (implicitly) assumed normal. From 1996q1 the Bank of England published the so-called "fan" chart, based on the two-piece normal distribution, that allows for skewness or the "balance of risks" to be on the upside or downside; see Britton et al. (1998). From 1997q3 these charts have been based on the deliberations of the Monetary Policy Committee (MPC).[14] The forecasts are then stated in the *Inflation Report* to "represent the MPC's best collective judgement about the most likely path for inflation... and the uncertainties surrounding those central projections". The measure of inflation targeted has changed over the sample period, from RPIX to CPI inflation, and we evaluate forecasts relative to the appropriate outturn.[15] Strictly the forecasts examined are conditional on unchanged interest rates.

The quality of the Bank of England's forecasts attracted considerable flak through the late 2000s and early 2010s. The Bank's Governor was forced to write several *mea culpa* letters to the U.K. Chancellor of the Exchequer as inflation persistently exceeded the target of 2% by more than 1% (as seen in the bottom panel of Figure 2). This repeated breach of its mandate from Parliament culminated in an official independent review - the Stockton Review - in 2012. This found the Bank's (point) forecasts to be

[14]The parameters of the density forecasts can be downloaded from the Bank of England's website.

[15]The final projection for RPIX inflation was published in the February 2004 *Inflation Report*.

"marginally" worse than outside Experts. It therefore seems appropriate to see whether Expert combination would have delivered improved (density) forecasts.

We consider combining the Bank of England (BoE) Expert density with a Climatological (or unconditional) Gaussian Expert. Such an Expert is popular in statistics as a reference forecast (e.g. see Gneiting et al. (2007)) and involves Least Squares projection of inflation on an intercept. With the mean inflation rate estimated to be just above 2% in our macroeconomic context this relates to the (two year ahead) inflation target which the BoE is charged with delivering.

Table 2 reports the average logarithmic scores of these two Experts and of combinations of their densities produced via the different opinion pools. The full-sample, T, is used to estimate the pools. We defer recursive (real-time) estimation of the parameters in the pools to future work, with higher T. The best performing pool's score is highlighted red.

Table 2 shows that relative forecast accuracy varies across the two Experts - BoE and Climatological - with the forecasting horizon, h. At $h = 1$, which is in fact a "nowcast", the BoE Expert is clearly better, but at $h = 8$ the Climatological Expert is preferred. Only at the medium range (around $h = 4$ to $h = 6$ quarters ahead) do we observe more equality in performance between the two Experts. This is critical in understanding when combination helps. From Table 2 we see that it is at these medium ranges that the copula pools do consistently deliver gains relative to the best individual Expert. The logarithmic score of the preferred pool is placed in red font. We see that from $h = 4$ to $h = 6$ this is the Symmetrized Joe-Clayton COP. The optimized linear opinion pool does beat the best individual Expert at $h = 4$ but at $h = 5$ to $h = 6$ returns the same density as the Climatological Expert, as the BoE Expert receives no weight in the combination. It is from $h = 4$ to $h = 6$ quarters ahead that the individual Experts' forecasts are also more dependent; the final row of Table 2 shows that the correlation between the conditional mean forecasts from the two Expert's densities is higher at these horizons than when $h = 1$ to $h = 3$ or $h = 7$ to $h = 8$.[16]

Consistent with the simulations, the benefits to the COP accrue when dependence between the Experts is higher. Table 2 shows that at the short horizons, when the BoE forecast is competitive, combination can in fact still help. But with the BoE Expert so superior to the Climatological one, the two Experts' forecasts are distinct; we observe a 0.0 correlation coefficient between their conditional mean forecasts at $h = 1$. This explains why the logOP, which assumes independence, works well. But since the COP nests the

[16]The correlation at $h = 4$ is higher than at $h = 7$ when we look to two decimal places.

Table 2: Combining the Bank of England and Climatological Experts: Average Logarithmic Score (1993q4-2011q4) by Forecast Horizon, h, and correlation (r) between the two Experts' pits

	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8
BoE	541	873	1149	1297	1387	1394	1773	2443
Climat.	1087	1077	1059	1047	1045	1046	1050	1053
norCOP	490	788	1002	1075	1074	1029	1116	1230
sjcCOP	838	1077	1172	1024	964	974	1412	2169
FrankCOP	490	796	1027	1129	1140	1092	1155	1269
tCOP	494	794	999	1062	1055	1019	1104	1230
LOPop	541	829	973	1028	1045	1046	1050	1053
LOP	684	852	975	1063	1110	1124	1144	1206
logOP	490	799	1050	1185	1237	1216	1224	1297
r	0.0	0.2	0.3	0.5	0.6	0.7	0.5	0.3

Notes: BoE and Climat. are the Bank of England and climatological (marginal) Expert forecast densities; norCOP, sjcCOP, tCOP and frankCOP are the copula opinion pools with normal, Symmetrized Joe-Clayton, t and Frank copula functions; LOPop is the linear opinion pool with optimized weights; LOP is the equal-weighted linear opinion pool; logOP is the logarithmic opinion pool.

logOP it is no surprise that the normal and Frank COPs deliver identical predictive densities and also minimize the logarithmic score at $h = 1$. At $h = 2$ and $h = 3$ dependence between the Expert densities increases and we do observe the COPs delivering more accurate densities than either Expert.

At the longer horizons, $h = 7$ and $h = 8$, the Climatological Expert is clearly preferred to the BoE Expert and the best combination is no combination at all. This is consistent with the optimized LOP returning a zero weight on the BoE Expert.

5.1 More detailed analysis of the one-year ahead forecasts

We now focus on the one-year ahead forecasts ($h = 5$). At this horizon the optimized LOP weight on the BoE Expert equals 0.03 explaining why the accuracy of the LOP matches that of the Climatological Expert. But this preference for a single Expert in the linear pool masks the considerable dependence that exists between the two Expert's density forecast errors as measured by their pits. Figure 1 plots the scatterplot between their pits and suggests these density forecasting errors do fan out, such that the two Experts have both made large negative density forecasting errors together but are surprised in different ways on the upside. This is confirmed when looking at the ML parameters from the estimated SJC-COP. $\widehat{\tau}^L = 0.76$ and $\widehat{\tau}^U = 0.15$ confirm lower tail dependence.

Figure 2 reveals that this lower tail dependence is consistent with both Experts making large negative density forecasting errors together over the recent recession. The top panel of Figure 2 shows that the quarter-by-quarter log scores of the sjcCOP, LOP and logOP all took a big hit over 2008-9 in the aftermath of the global financial crisis. But explicitly assuming Expert independence, as in the logOP, is the worst idea, with the lowest scores returned over the recession. The optimized LOP, which recall is essentially the Climatological Expert, in fact forecasts best over the recession itself, but this is at the expense of consistently lower scores than the sjcCOP before and after the recession. This is because, as the middle panel of Figure 2 indicates, the optimized LOP delivers a far too wide density forecast, explaining why the probability of inflation exceeding 3% is essentially unchanged from 1993-2011. By contrast, the sjcCOP exhibits more variation, reflecting its sharper densities. Comparison with the bottom panel of Figure 2 indicates that the sjcCOP does correctly anticipate the rise in inflation above the 3 per cent target from 2010.

The superiority of these probability event forecasts constructed from the three pools is confirmed statistically when we undertake encompassing tests following Clements &

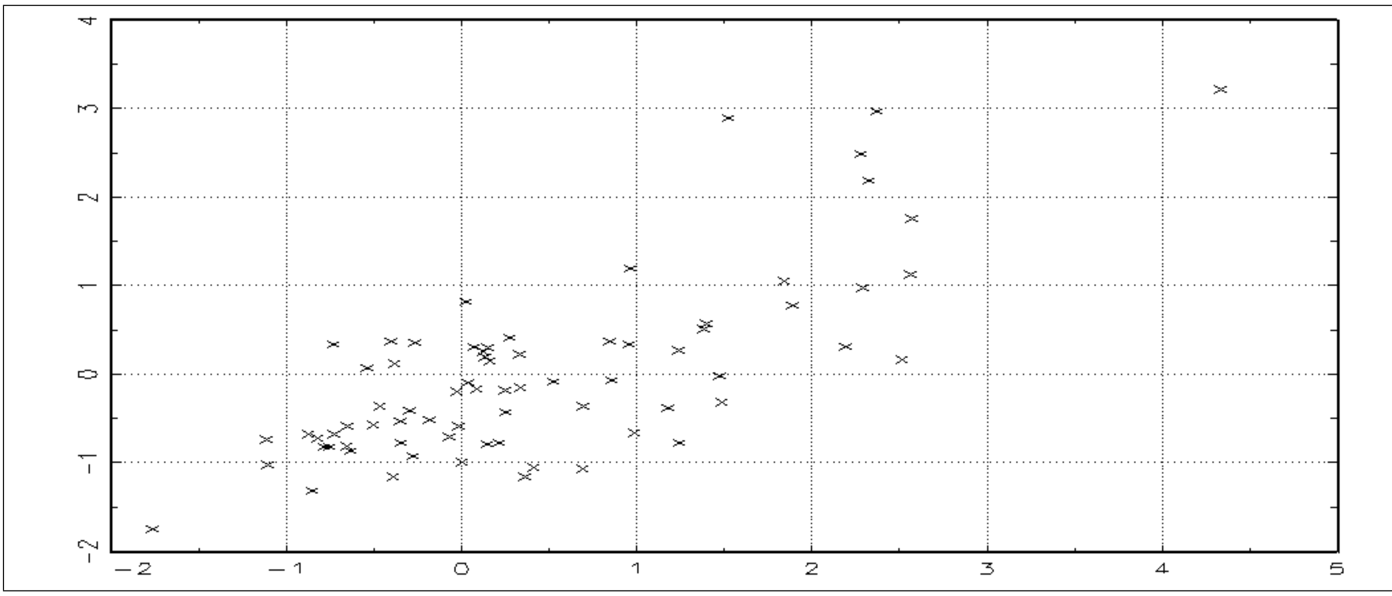

Figure 1: Bivariate inverse normal *pits* plot between the Bank of England and Climatological Experts at $h = 5$

Harvey (2010). These involve using a logit model to relate the binary outturn (was inflation greater than 3% or not) to the three probability event forecasts. We find t-ratios of 3.1 on the sjcCOP, 1.5 on the LOP and -1.7 on the logOP. One also cannot reject the null hypothesis that the sjcCOP *encompasses* the other two Experts at a p-value of 2%.

We conclude the empirical application by stressing the importance of the model space used as the basis for combination or pooling. As discussed above, the performance of the different opinion pools is sensitive to the quality and relationship between the component Expert densities. If one Expert density dominates the other(s), it should be no surprise when no pool helps. As recent econometric work has found combining Expert densities is beneficial when there is uncertainty about the preferred Expert (e.g., see Jore et al. (2010)) and when the Decision Maker suspects that all Expert densities are likely incorrect (see Geweke & Amisano (2011)).

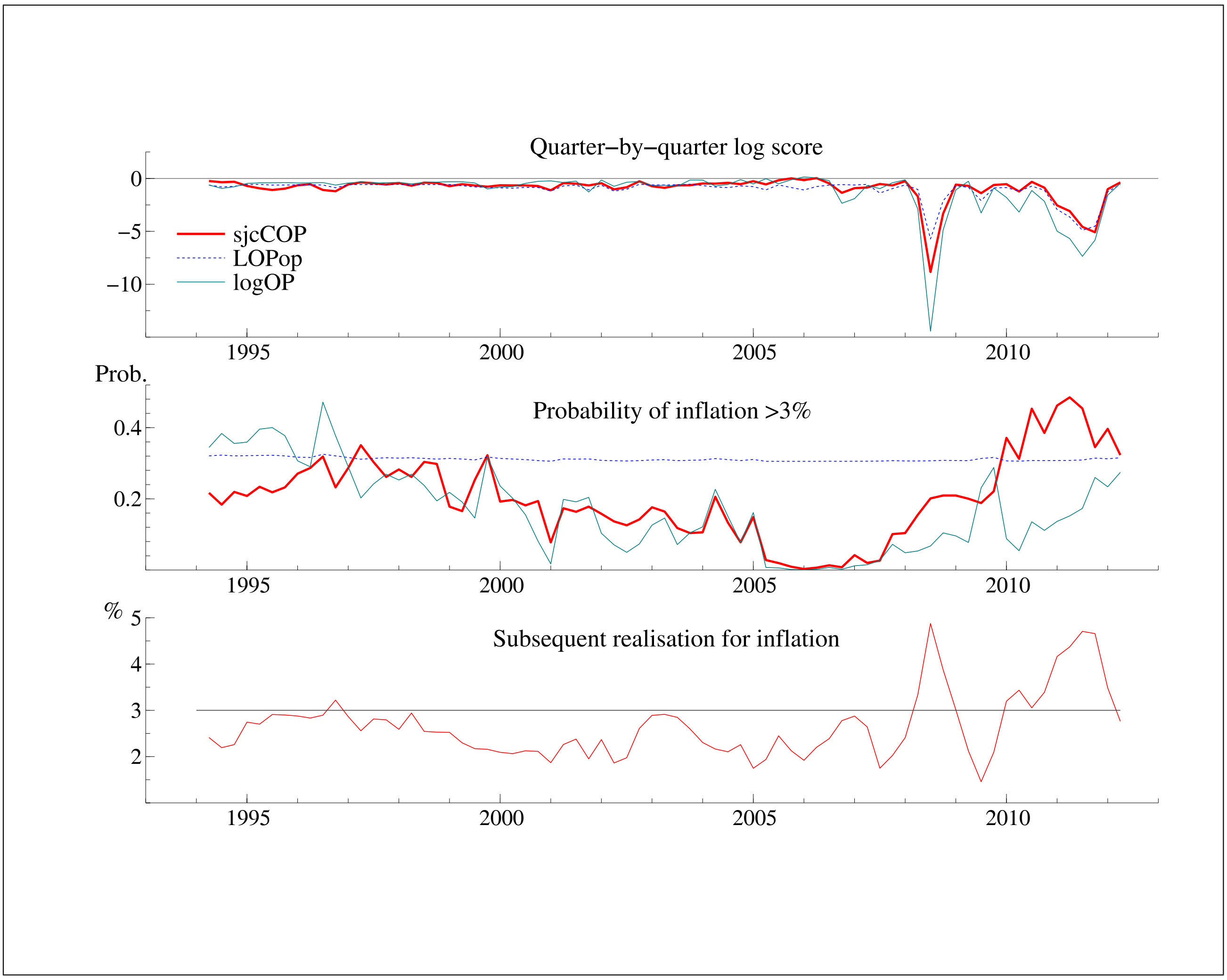

Figure 2: One year ahead probability event forecasts for the Symmetrized Joe-Clayton COP, the optimized LOP and the logarithmic opinion pool

6 Conclusion

This paper sets out a general framework to accommodate stochastic dependence, including asymmetric dependence, between multiple Experts' probability density forecasts. In so doing it proposes a recalibrated and copula opinion pool with focus on the latter. Simulations indicate the flexibility and superiority of the proposed copula opinion pool and an application using forecasts from the Bank of England reveals its utility to Decision Makers when combining dependent Expert densities. The proposed opinion pools, like performance weighted and optimized linear pools, are designed for use in situations when a Decision Maker is confronted by a fixed set of competing Expert densities and uses historical data to maximize the fit of the pool. In our case this involves using historical data, perhaps simulated for statistical Experts, to determine the degree and type of dependence, as captured by the copula function, between the Experts' densities. We should expect (and indeed find) non-zero dependence between Experts' densities in many situations, and modeling this via the proposed copula opinion pool therefore provides scope for improved combined density forecasts.

Future work should consider applications with more than two Experts, where there may be gains to trading off flexibility with parsimony when specifying higher-dimensional copula functions. And examine performance out-of-sample. In addition, the proposed recalibrated and copula opinion pools might be compared with linear and logarithmic alternatives, that ignore dependence, using analytic methods of the sort proposed by Hora (2010). This would involve extending Hora (2010) to consider more general forms of dependence than (Pearson) correlation. Thereby, one could further study how Expert dependence affects the properties of competing opinion pools. The results in this paper suggest that the proposed recalibrated and copula opinion pools offer a promising yet practical means of improving forecast accuracy in the face of general forms of Expert dependence.

References

Bates, J. M. & Granger, C. W. J. (1969), 'The combination of forecasts', *Operational Research Quarterly* **20**, 451–468.

Britton, E., Fisher, P. & Whitley, J. (1998), 'The inflation report projections: understanding the fan chart', *Bank of England Quarterly Bulletin* **38**, 30–37.

Chen, X., Fan, Y. & Patton, A. J. (2004), Simple tests for models of dependence between multiple financial time series, with applications to U.S. equity returns and exchange rates. L.S.E. Discussion Paper 483.

Clemen, R. (1986), 'Calibration and the aggregation of probabilities', *Management Science* **32**, 312–314.

Clemen, R. T. & Reilly, T. (1999), 'Correlations and copulas for decision and risk analysis', *Management Science* **45**, 208–224.

Clemen, R. & Winkler, R. (1999), 'Combining probability distributions from experts in risk analysis', *Risk Analysis* **19**, 187–203.

Clements, M. P. & Harvey, D. I. (2010), 'Forecast encompassing tests and probability forecasts', *Journal of Applied Econometrics* **25**(6), 1028–1062.

Clements, M. P. & Hendry, D. F. (1998), *Forecasting Economic Time Series*, Cambridge University Press: Cambridge.

Cooke, R. (1991), *Experts in Uncertainty: Opinion and Subjective Probability in Science*, New York: Oxford University Press.

Cooke, R. & Goossens, L. (2008), 'TU Delft expert judgment data base', *Reliability Engineering and System Safety* **93**, 657–674.

Embrechts, P., McNeil, A. & Straumann, D. (1999), Correlation and dependence properties in risk management: properties and pitfalls, *in* M. Dempster, ed., 'Risk Management: Value at Risk and Beyond', Cambridge: Cambridge University Press.

Garratt, A., Mitchell, J., Vahey, S. P. & Wakerly, E. C. (2011), 'Real-time inflation forecast densities from ensemble phillips curves', *North American Journal of Economics and Finance* **22**, 77–87.

Genest, C. & Zidek, J. (1986), 'Combining probability distributions: a critique and an annotated bibliography', *Statistical Science* **1**, 114–135.

Geweke, J. & Amisano, G. (2011), 'Optimal prediction pools', *Journal of Econometrics* **164**(1), 130–141.

Giacomini, R. & White, H. (2006), 'Tests of conditional predictive ability', *Econometrica* **74**, 1545–1578.

Gneiting, T., Balabdaoui, F. & Raftery, A. E. (2007), 'Probabilistic forecasts, calibration and sharpness', *Journal of the Royal Statistical Society: Series B* **69**, 243–268.

Gneiting, T. & Raftery, A. E. (2007), 'Strictly proper scoring rules, prediction, and estimation', *Journal of the American Statistical Association* **102**, 359–378.

Hall, S. G. & Mitchell, J. (2007), 'Combining density forecasts', *International Journal of Forecasting* **23**, 1–13.

Halperin, M. (1961), 'Almost linearly-optimum combination of unbiased estimates', *Journal of the American Statistical Association* **56**, 36–43.

Hora, S. (2010), 'An analytic method for evaluating the performance of aggregation rules for probability densities', *Operations Research* **58**, 1440–1449.

Joe, H. (1997), *Multivariate models and dependence concepts*, London: Chapman and Hall.

Jore, A. S., Mitchell, J. & Vahey, S. P. (2010), 'Combining forecast densities from VARs with uncertain instabilities', *Journal of Applied Econometrics* **25**, 621–634.

Jouini, M. N. & Clemen, R. T. (1996), 'Copula models for aggregating expert opinions', *Operations Research* **44**, 444–457.

Kapetanios, G., Labhard, V. & Price, S. (2008), 'Forecast combination and the bank of england's suite of statistical forecasting models', *Economic Modelling* **25**(4), 772–792.

Lindley, D. (1983), 'Reconciliation of probability distributions', *Operations Research* **31**, 866–880.

Mitchell, J. & Hall, S. G. (2005), 'Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESR "fan" charts of inflation', *Oxford Bulletin of Economics and Statistics* **67**, 995–1033.

Mitchell, J. & Wallis, K. (2011), 'Evaluating density forecasts: Forecast combinations, model mixtures, calibration and sharpness', *Journal of Applied Econometrics* **26**, 1023–1040.

Morris, P. (1974), 'Decision analysis expert use', *Management Science* **20**, 1233–1241.

Morris, P. (1977), 'Combining expert judgments: A Bayesian approach', *Management Science* **23**, 679–693.

Nelsen, R. B. (1999), *An introduction to copulas*, Lecture Notes in Statistics 139: Springer.

Patton, A. J. (2006), 'Modelling asymmetric exchange rate dependence', *International Economic Review* **47**, 527–556.

Raftery, A. E., Gneiting, T., Balabdaoui, F. & Polakowski, M. (1995), 'Using Bayesian Model Averaging to Calibrate Forecast Ensembles', *Monthly Weather Review* **133**, 1155–1174.

Ranjan, R. & Gneiting, T. (2010), 'Combining probability forecasts', *Journal of the Royal Statistical Society Series B* **72**(1), 71–91.

Sklar, A. (1959), 'Fonctions de répartition à n dimensions et leurs marges', *Publications de l'Institut Statistique de l'Université de Paris 8* pp. 229–31.

Smith, J. & Wallis, K. F. (2009), 'A simple explanation of the forecast combination puzzle', *Oxford Bulletin of Economics and Statistics* **71**, 331–355.

Winkler, R. (1981), 'Combining probability distributions from dependent information sources', *Management Science* **27**, 479–488.

Winkler, R. (1986), 'Expert resolution', *Management Science* **32**, 298–303.

Zellner, A. (1986), 'Biased predictors, rationality and the evaluation of forecasts', *Economics Letters* **21**, 45–48.