Predictive Densities for Shire Level Wheat Yield in Western Australia

William E. Griffiths, Lisa S. Newton and Christopher J. O'Donnell

University of New England

Rainfall during the germination, growing and flowering periods is a major determinant of wheat yield. The degree of uncertainty attached to a wheat-yield prediction depends on whether the prediction is made before or after the rainfall in each period has been realised. Bayesian predictive densities that reflect the different levels of uncertainty in wheat-yield predictions made at four different points in time are derived for five shires in Western Australia.

Key Words: Bayesian prediction; Prediction with uncertain regressors.

Paper contributed to the Australian Agricultural and Resource Economics Society Conference, Adelaide, 23-25 January, 2001.

1. Introduction

Shire level forecasts of wheat yield provide useful information for those involved in wheat transportation, storage and marketing. In the absence of micro-level data on soil moisture, the nutrient content of the soil, the presence of weeds and pests, etc., a reasonable method for modelling shire-level wheat yield is via an equation that relates yield to a trend term (to accommodate technological change), and to rainfall at different times during the year (Coelli 1992 and references therein). In what follows, we consider three rainfall periods, germination (May and June), development (July and August) and flowering (September and October). We are concerned with forecasting wheat yields before and after each of these periods. When a rainfall variable is an explanatory variable in a regression equation designed for forecasting yield, having to forecast before that variable has been observed raises a methodological problem. It is this methodological problem that is the main focus of this paper.

Suppose that a yield forecast is to be made after rainfall at flowering has been observed. Since the other rainfall periods precede the flowering period, this forecast is made with complete knowledge of all regressors. Conventional regression forecasting methodology is applicable. However, making a forecast prior to the flowering period means that uncertainty about rainfall during that period must be recognised in the forecasting procedure. Making a forecast prior to the development period means that rainfall uncertainty in two periods must be recognised. Similarly, a forecast prior to germination means that rainfall uncertainty in all three periods must be captured. The four possible forecast times and their relationship to the rainfall periods are depicted in Figure 1. We are concerned with the development of techniques that recognise the different levels of uncertainty associated with the timing of the forecast.

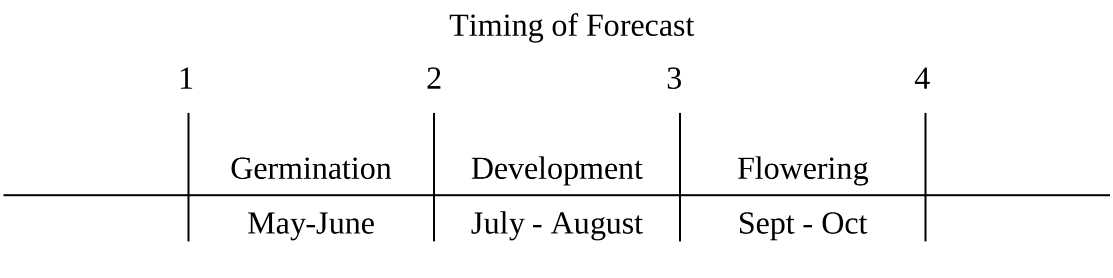


Figure 1: Forecast Times and Rainfall Periods

The forecasting tool that we employ in conjunction with the linear regression model is the Bayesian predictive probability density function (predictive pdf). This tool has the advantage of providing a unified approach that conditions only on past data and the chosen model. Uncertainty about regression coefficients, about unknown future rainfall, and about the unknown parameters of the rainfall distribution, are all captured in the predictive pdf. Distinctions between forecasts made after and before rainfall has been observed are readily made by conditioning, or not conditioning, on that rainfall. Another advantage of the Bayesian approach is the scope for improving predictions by including prior information on the regression coefficients. In the shires that we consider, it is reasonable to insist that the response of yield to rainfall, evaluated at the sample mean of rainfall, is positive. Inequality restrictions of this kind can be readily incorporated into the analysis.

The literature on crop yield forecasting is vast, ranging from studies that use micro-level data from experimental plots to those, like our study, that use more limited aggregate data. For an extensive review, see Stephens (1995); for a more restrictive review, but one with more focused relevance to what follows in our paper, see Coelli (1992). In contrast, the literature on regression-model forecasting with uncertain regressors is somewhat limited. Some efforts have been made in the time-series arena (e.g., Baillie 1979). Recognising an inconsistency described by Feldstein (1971), McCullough (1996) suggests using the bootstrap to obtain consistent forecast intervals. Various approaches to obtaining forecast intervals for a range of models have been reviewed by Chatfield (1993). In this paper we illustrate how Bayesian principles can be readily applied to obtain forecasts and forecast intervals that reflect uncertainty from unknown regressors and the consequent uncertainty from the timing of the forecast.

In Section 2 we describe the model and data. Some notation is introduced in Section 3, and the predictive pdf's that are the subject of the paper are discussed. Section 4 is concerned with modelling rainfall; the model is described, the estimation-prediction procedure is outlined, and results for the five shires are presented. In Section 5 we describe how to obtain posterior pdf's for the regression coefficients; corresponding results are presented. The results obtained in Sections 4 and 5 are drawn together in Section 6 and utilised for forecasting. This utilisation requires some theory to unify the earlier parts; the forecasting results are then presented. Some concluding remarks are made in Section 7.

2. Model and Data

To illustrate the methodology, five shires are chosen from the northern part of the Western Australian wheat belt: Northampton, Chapman Valley, Mullewa, Greenough and Irwin. The paper by Coelli (1992) motivated the methodology that we develop; the chosen shires are a subset of those that he considered. Extensions of our methodology to allow for hierarchical priors and spatial autocorrelation are currently being considered by Newton (2001). Rainfall data were obtained from the Western Australian office of the Bureau of Meteorology. The rainfall for a given shire was taken as the measured rainfall at a site considered representative of that shire. These sites were Northampton P.O. (for Northampton shire), Chapman Research Station at Nabawa (for Chapman Valley shire), Mullewa (for Mullewa shire), Geraldton airport (for Greenough shire), and Dongara (for Irwin shire). Data on area of wheat planted and production of wheat by shire were obtained from Coelli (1992) for the early years and from the Australian Bureau of Statistics for the later years. Yield data were obtained by dividing production by area. A total of 47 observations were used for estimation, covering the period 1950-1996. The forecasting problem is to predict 1997 yield in each of the five shires. Forecasts are made with, without, and with partial information on 1997 rainfall data. Realized yields in 1997 are compared with forecasts.

To explain wheat yield in year *t*, denoted by Y_t, the following model was specified for each of the shires

$$Y_t = \beta_1 + \beta_2 T_t + \beta_3 T_t^{(2)} + \beta_4 T_t^{(3)} + \beta_5 G_t + \beta_6 G_t^2 + \beta_7 D_t + \beta_8 D_t^2 + \beta_9 F_t + \beta_{10} F_t^2 + e_t \tag{2.1}$$

It was assumed that the error term e_t was distributed as independent $N(0,\sigma^2)$; tests for autocorrelation, heteroskedasticity and non-normality did not provide evidence to the contrary. The terms T_t, $T_t^{(2)}$ and $T_t^{(3)}$ describe a cubic trend designed to capture technological change, such as improved varieties and changes in farming practice. The inclusion of a trend for this purpose is common practice; Coelli (1992) cites several examples. A cubic trend was chosen on the basis of goodness-of-fit, after experimenting with other trends including linear, quadratic and exponential. Time was indexed as $t = 1, 2, ..., 47$ and the trend terms were scaled as $T_t = t/1000$, $T_t^{(2)} = t^2/1000$ and $T_t^{(3)} = t^3/1000$.

The variables G_t, D_t and F_t are rainfall variables, referring to rainfall in the germination, development and flowering periods, respectively. They are expressed as ratios, relative to average rainfall over the sample period. As mentioned in the introduction, germination period rainfall is taken as that rainfall during May and June, the development period refers to July and August and flowering refers to September and October. Yield is specified as a quadratic function of rainfall in each of the three periods to allow for a decreasing or an increasing marginal yield.

3. Notation and Predictive pdf's

Before developing the required methodology, it is convenient to introduce some notation. Because each shire is treated separately, the same notation is used to refer to any of the five shires.

$Y = \{Y_t \mid t = 1, 2, ..., 47\} =$ observed yields over the sample period 1950-1996.

$Y^* = Y_{48} =$ the 1997 yield value that is being forecast.

$X = \{(G_t, D_t, F_t) \mid t = 1, 2, ..., 47\} =$ observed rainfalls over the sample period.

$G^* = G_{48} =$ rainfall during the 1997 germination period.

$D^* = D_{48} =$ rainfall during the 1997 development period.

$F^* = F_{48} =$ rainfall during the 1997 flowering period.

$\beta = (\beta_1, \beta_2, ..., \beta_{10})' =$ the vector of unknown regression coeficients.

Bayesian inference, whether it be for estimation or forecasting, is concerned with the derivation and estimation of conditional probability density functions (pdf's). In some of the conditional pdf's that we utilise, it is convenient to economise on notation. Since the trend terms are always nonstochastic and known, they will not be written explicitly as conditioning variables. Where required, their presence is implicit. Also, since knowledge of G^* (say) implies knowledge of G^{*2}, derivation of a pdf for G^*, or conditioning on G^*, will be regarded as sufficient; explicit recognition of G^{*2} is not necessary.

The Bayesian predictive pdf for a yield summarises what we know about possible future values of that yield and their likelihood of being realized. Because we are interested in predicting yields with and without knowledge of the rainfall variables G^*, D^* and F^*, different predictive pdf's are required for the predictions made at different points in time. Using $f(.)$ as generic notation for a pdf, the predictive pdf's of interest are:

$f(Y^* | Y, X, G^*, D^*, F^*) =$ the predictive pdf for yield after rainfall at flowering has been observed,

$f(Y^* | Y, X, G^*, D^*) =$ the predictive pdf for yield at the end of August (before flowering rainfall has been observed, but after development rainfall),

$f(Y^* | Y, X, G^*) =$ the predictive pdf for yield at the end of June (before development rainfall has been observed, but after germination rainfall),

$f(Y^* | Y, X) =$ the predictive pdf for yield before germination rainfall has been observed.

Note that each of these pdf's is conditional on the sample values of yield (Y) and rainfall (X), and on the observed rainfalls in the forecast year. They are not conditional on the unobserved rainfalls in the forecast year, nor the unknown parameters (β, σ). When forecasting, the consequence of uncertainty about unobservables is captured by a predictive pdf that is not conditional on these unobservables.

Each of the four predictive pdf's is obtained for each of the five shires, with and without imposing inequality constraints on the regression coefficients. Thus, a total of $4 \times 5 \times 2 = 40$ predictive pdf's are computed. To reduce presentation of the results to manageable proportions, complete predictive pdf's are graphed for only selected cases. Means and standard deviations of the predictive pdf's are tabulated for all cases. Proceeding with and without the inequality restrictions gives us the opportunity to assess the effect of prior information that insists the response of yield to rainfall (at mean rainfall) is positive.

4. Modelling Rainfall

For forecasting yield prior to realization of rainfall in one or more of the 1997 rainfall periods, it is necessary to have a model for predicting rainfall. In what follows we consider predicting G^*; similar methodology is used for predicting rainfall in the development and flowering periods. We treat G^*, D^* and F^* separately because tests for independence did not reveal any significant correlations between the rainfalls in different periods. Nor was there any significant year-to-year autocorrelations.

We assumed that rainfall in a given period follows a truncated normal distribution, truncated from below by zero. Chi-square goodness-of-fit tests on the sample observations did not suggest this assumption was unrealistic. Let $T = 47$ and let $G = (G_1, G_2, ..., G_{47})'$. Denote the location and scale parameters of the truncated normal distribution as μ and τ, respectively. The joint pdf for G is

$$f(G\,|\,\mu,\tau) = (2\pi)^{-T/2}\left[\Phi\left(\frac{\mu}{\tau}\right)\right]^{-T} \tau^{-T} \exp\left\{-\frac{1}{2\tau^2}\sum_{t=1}^{T}(G_t-\mu)^2\right\} \tag{4.1}$$

where $\Phi(.)$ is the standard normal cumulative distribution function. The predictive pdf for G^*, that becomes important for forecasting yield when G^* is unknown, is given by

$$\begin{aligned} f(G^*\,|\,G) &= \iint f(G^*,\mu,\tau\,|\,G)\, d\mu\, d\tau \\ &= \iint f(G^*\,|\,\mu,\tau)\, f(\mu,\tau\,|\,G)\, d\mu\, d\tau \end{aligned} \tag{4.2}$$

In this expression $f(\mu,\tau\,|\,G)$ is the posterior pdf for (μ,τ). The pdf $f(G^*\,|\,\mu,\tau)$ is the truncated normal distribution; it is not necessary for this latter pdf to be conditioned on G because rainfall is independent from year to year.

Generally (and certainly in our case), the integrals that define $f(G^* | G)$ are intractable. However, we can proceed via simulation. What turns out to be important for forecasting yield is that we are able to draw observations on G^* from the predictive pdf $f(G^* | G)$. We can do so by first drawing values (μ,τ) from their posterior pdf $f(\mu,\tau | G)$; then for each pair (μ,τ), a value of G^* is drawn from $f(G^* | \mu,\tau)$. Simple acceptance-rejection sampling is satisfactory for drawing from the truncated normal distribution $f(G^* | \mu,\tau)$. That is, draws are made from a normal distribution that is not truncated and, if a negative value for G^* is obtained, it is discarded and a replacement draw is made. To obtain the posterior pdf for (μ,τ), we need to first specify a prior pdf and to then apply Bayes' theorem. As a prior, we use the conventional noninformative prior $f(\mu,\tau) \propto \tau^{-1}$. Application of Bayes' theorem yields the posterior pdf

$$\begin{aligned} f(\mu,\tau | G) &\propto f(G | \mu,\tau)\, f(\mu,\tau) \\ &\propto \left[\Phi\left(\frac{\mu}{\tau}\right)\right]^{-T} \tau^{-(T+1)} \exp\left\{-\frac{1}{2\tau^2}\sum_{t=1}^{T}(G_t - \mu)^2\right\} \end{aligned} \tag{4.3}$$

Because this pdf is not of a recognisable tractable form, we use a Metropolis-Hastings algorithm to draw observations from it. Details are in an appendix. A total of 50,000 draws were made, with 5,000 being discarded as a burn-in. For each of the retained 45,000 draws of (μ,τ), a draw of G^* was made. How these draws are used to forecast yield is discussed in Section 6.

The posterior means and standard deviations for (μ,τ) as well as the means and standard deviations of the predictive pdfs for G^*, D^* and F^*, for each of the five shires, are given in Table 1. These values are estimates obtained from the 45,000 draws. With the exception of flowering rainfall for Mullewa, the posterior standard deviations for μ and τ are all relatively small, indicating that we are accurately estimating these parameters. The truncation of the distribution at zero does have an effect, but not a severe one. The magnitude of the truncation can be assessed by comparing the posterior mean of τ relative to that of μ, and by examining the amount by which the predictive mean for rainfall exceeds the posterior mean for μ.

The variation in rainfall is similar for all shires; flowering rainfall is more variable than that during germination and development.

5. Estimating Regression Parameters

5.1 Without prior inequality restrictions

Two sets of posterior pdf's for the regression parameters (β,σ) were obtained, one using a noninformative prior pdf and one using a prior pdf that includes inequality restrictions. The noninformative prior pdf that we employ is the conventional one $f(\beta,\tau)\propto\sigma^{-1}$. It is well known (see, for example Judge et al 1988, p.318) that the marginal posterior pdf for β, obtained after applying Bayes' theorem and then integrating out σ is

$$f(\beta\,|\,Y,X)\propto\left[1+(\beta-b)'\frac{Z'Z}{vs^2}(\beta-b)\right]^{-(K+v)/2} \tag{5.1}$$

In this expression, Z is the (47×10) matrix that contains the sample observations on the constant, the time trend, the rainfall variables and their squares. We are using X to denote the set of rainfalls in the sample period and Z to denote the complete regressor matrix. The terms b and s^2 refer to the least squares estimator $b=(Z'Z)^{-1}Z'Y$ and the error variance estimator $s^2=(Y-Zb)'(Y-Zb)/v$, where $v=T-K$, and $K=10$ is the dimension of β. The pdf $f(\beta\,|\,Y,X)$ is a multivariate *t*- distribution with mean b and covariance matrix $(v/(v-2))s^2(Z'Z)^{-1}$. It is the elements in b and the square roots of the diagonal elements of $(v/(v-2))s^2(Z'Z)^{-1}$ that provide the posterior means and standard deviations reported in Table 2.

The marginal posterior pdf for σ, obtained by applying Bayes' theorem and then integrating out β is

$$f(\sigma\,|\,Y,X)\propto\frac{1}{\sigma^{v+1}}\exp\left\{-\frac{vs^2}{2\sigma^2}\right\} \tag{5.2}$$

This pdf is an "inverted gamma" distribution (see Zellner 1971, p.371); its posterior mean and standard deviation are $E(\sigma\,|\,Y,X)=\left(\Gamma[(v-1)/2]/\Gamma(v/2)\right)\left(v/2\right)^{1/2}s$ and

$[(v/(v-2))s^2 - (E(\sigma|Y,X))^2]^{1/2}$, respectively. These values are also reported in Table 2.

From Table 2 we observe that:

1. The trends for all shires have similar shapes although there is some variation in the coefficients of the trend terms across shires. An example of the shape of the trend is given in Figure 2 for Greenough shire.

2. Using the posterior means as estimates, the yield response to rainfall is substantially different for the five shires, despite the fact that all shires have comparable average rainfalls with the exception of Mullewa whose rainfall is lower.

3. With the exception of the flowering period in Greenough and the development period in Irwin, the posterior means exhibit a decreasing marginal response of yield to rainfall for all rainfall periods in all shires. In Greenough's flowering period there is an increasing marginal response. In Irwin's development period we have the counter-intuitive outcome of a negative response of yield to rainfall except at very high rainfalls. The posterior standard deviations of the offending coefficients are relatively high, however, and so the existence of more realistic outcomes is not precluded.

Further remarks will be made after introducing the inequality restrictions.

5.2 With prior inequality restrictions

To introduce the inequality restrictions on yield response to rainfall, we begin by noting that the response of yield to germination rainfall, for example, is

$$\frac{\partial Y}{\partial G} = \beta_5 + 2\beta_6 G$$

By construction, the sample mean of G is one. Thus, the inequality restriction that yield response to rainfall must be positive at mean sample rainfall can be represented as $\beta_5 + 2\beta_6 > 0$. Using a similar argument for the other rainfall periods, we can define a feasible parameter region as

$$R(\beta) = \{\beta \,|\, \beta_5 + 2\beta_6 > 0,\ \beta_7 + 2\beta_8 > 0,\ \beta_9 + 2\beta_{10} > 0\} \tag{5.3}$$

Furthermore, let $I_R(\beta)$ be an indicator function that equals 1 when $\beta \in R(\beta)$ and 0 otherwise. Then a prior pdf for (β, σ) that restricts β to the feasible region, but is otherwise noninformative is given by

$$f(\beta, \sigma \mid R) \propto \frac{I_R(\beta)}{\sigma} \tag{5.4}$$

Applying Bayes' theorem and integrating σ out of the joint posterior pdf, gives the marginal posterior pdf for β

$$f(\beta \mid Y, X, R) \propto \left[1 + (\beta - b)' \frac{Z'Z}{vs^2}(\beta - b)\right]^{-(K+v)/2} I_R(\beta) \tag{5.5}$$

To distinguish this case from that without the inequality constraints, we have included R as a conditioning "variable" in (5.4) and (5.5). The pdf in (5.5) is a truncated multivariate *t*-distribution. Using it to find marginal posterior pdf's for each of the single parameters in β, and their posterior means and standard deviations, is not straightforward. Also, because it involves multiple inequality restrictions, sampling from it using a simple acceptance-rejection algorithm is not efficient. We use a Metropolis-Hastings algorithm instead. See the appendix. After discarding 5,000 draws for a burn-in, 45,000 draws are retained for analysis. The posterior means and standard deviations reported in Table 3 are estimates from these 45,000 draws.

The marginal posterior pdf for σ can no longer be written in the inverted-gamma form that was given in equation (5.2). To appreciate why, note that

$$f(\beta, \sigma \mid Y, X, R) = f(\beta \mid \sigma, Y, X, R)\ f(\sigma \mid Y, X, R) \tag{5.6}$$

It is possible to show that the conditional posterior pdf $f(\beta \mid \sigma, Y, X, R)$ is a truncated normal distribution. Because it is truncated, its normalising constant includes an unpleasant multivariate normal integral. This integral is a function of σ. If the joint posterior pdf $f(\beta, \sigma \mid Y, X, R)$ is separated into its conditional and marginal components as in equation (5.6), then the truncation integral needs to be included in the denominator of $f(\beta \mid \sigma, Y, X, R)$. Correspondingly, it appears in the numerator of $f(\sigma \mid Y, X, R)$, implying that this pdf is no longer an inverted gamma. We consider instead the conditional posterior pdf $f(\sigma \mid \beta, Y, X, R)$. It can be written as

$$f(\sigma \mid \beta, Y, X, R) \propto \frac{1}{\sigma^{T+1}} \exp\left\{-\frac{(Y - Z\beta)'(Y - Z\beta)}{2\sigma^2}\right\} \tag{5.7}$$

Like equation (5.2), this pdf is an inverted-gamma pdf. For each value of β drawn from the truncated *t* pdf in equation (5.5), we can draw a value of σ from (5.7). Proceeding in this way is equivalent to drawing σ from its marginal pdf. The posterior means and standard deviations reported in Table 3 are estimated from draws obtained in this way.

Before discussing the results in Table 3, it is instructive to re-examine Table 2 in light of the inequality restrictions. In five out of the fifteen rainfall period/shire combinations, the posterior means violate the restrictions. However, the violations are not so severe that they cast doubt on the validity of the inequalities. Given the noninformative prior, we can use the *t*-distribution in (5.1) to compute the probability that an inequality restriction holds. These probabilities are presented in Table 4. The five cases where the probability is less than 0.5 correspond to the cases where the posterior means violate the restrictions. If we are prepared to believe the inequality restrictions must hold, then the scope for improving estimation is considerable. The restrictions will have an impact not only when those posterior means from the noninformative prior violate the restrictions; the results will change as long as the probabilities in Table 4 are less than unity. Nevertheless, the largest effects of the restrictions do occur in the periods with posterior-mean violations. In Table 3 the posterior means from the inequality-restricted prior differ most from those in Table 2 for the germination and development periods in Northampton and Irwin, and the germination period in Greenough. The development period in Irwin now exhibits positive and decreasing marginal response to rainfall at the mean-rainfall point. Overall, the introduction of the inequality restrictions has tended to lower the posterior standard deviations, but this change is not universal. The small changes in the posterior means for σ are one indication of a lack of conflict between the prior information and the information from the data.

6. Forecasting

To develop the forecasting methodology, it is useful to introduce notation to distinguish between 1997 rainfalls that have been observed at the time the forecast is being made, and those that have not. We will use x for observed rainfalls and x^* for unobserved rainfalls. The elements in and dimensions of these vectors vary depending on the time the forecast is made. Thus, a single representation of the four predictive pdf's listed in Section 3 is $f(Y^* | Y, X, x)$. A theoretical representation of how this pdf is obtained is

$$f(Y^* | Y, X, x) = \int\int\int\int\int f(Y^*, \beta, \sigma, x^*, \mu, \tau | Y, X, x)\, d\beta\, d\sigma\, dx^*\, d\mu\, d\tau \qquad (6.1)$$

The definitions of μ and τ are broader in this equation. They need to be viewed as vectors with dimension equal to the number of unobserved rainfall components. The joint pdf on the right hand side of equation (6.1) involves all unknown quantities (future yield, regression parameters, future rainfalls, the parameters of the rainfall distribution), and conditions on known quantities (yields and rainfalls during the sample period, observed 1997 rainfall). The marginal pdf on the left hand side of the equation is obtained by integrating out of the joint pdf the unknown quantities that are not of direct interest. By integrating out these unknowns, rather than conditioning on them, or estimates of them, we are obtaining a marginal predictive pdf for yield that reflects uncertainty in all unknown quantities. It is not possible to evaluate every integral in (6.1) analytically. However, a mix of analytical integration and estimation of integrals is possible. The draws of x^*, obtained in Section 4, and the draws of (β, σ) obtained in Section 5 are useful for estimation.

6.1 With known rainfalls and no inequality restrictions

Not surprisingly, when 1997 rainfall in every period has been observed, equation (6.1) simplifies considerably. It becomes

$$f(Y^* | Y, X, G^*, D^*, F^*) = \int\int f(Y^*, \beta, \sigma | Y, X, G^*, D^*, F^*)\, d\beta\, d\sigma \qquad (6.2)$$

This predictive pdf makes provision for uncertainty in (β, σ); no provision for rainfall uncertainty is necessary. The next step depends on whether or not the inequality restrictions have been imposed. Without inequality restrictions analytical integration is possible yielding (see, for example, Zellner 1971, p.72)

$$f(Y^* | Y, X, G^*, D^*, F^*) \propto \left[1 + \frac{(Y^* - zb)'(Y^* - zb)}{vs^2 (1 + z(Z'Z)^{-1} z')} \right]^{-(v+1)/2} \qquad (6.3)$$

where z is a (1×10) row vector containing the constant, the trend terms for 1997, and the observed rainfalls and their squares. This pdf is a t-distribution with v degrees of freedom and mean and variance given by

$$E(Y^* \mid Y, X, G^*, D^*, F^*) = zb \tag{6.4}$$

$$\text{var}(Y^* \mid Y, X, G^*, D^*, F^*) = \left(\frac{v}{v-2}\right) s^2 (1 + z(Z'Z)^{-1} z') \tag{6.5}$$

6.2 With known rainfalls and inequality restrictions

After introducing the inequality constraints, and conditioning on R to make introduction of the inequality region explicit, we obtain $f(Y^* \mid Y, X, G^*, D^*, F^*, R)$ by estimating the integral in (6.2). The mean and variance of the predictive pdf are also estimated. To describe this estimation procedure, we first note that

$$f(Y^*, \beta, \sigma \mid Y, X, G^*, D^*, F^*, R) = f(Y^* \mid \beta, \sigma, G^*, D^*, F^*)\; f(\beta, \sigma \mid Y, X, R)$$

In this and later equations, where we write a joint pdf as the product of conditional and marginal pdf's, we omit conditioning variables that are redundant. Now, the conditional pdf $f(Y^* \mid \beta, \sigma, G^*, D^*, F^*)$ is a normal distribution with mean

$$E(Y^* \mid \beta, \sigma, G^*, D^*, F^*) = z\beta \tag{6.6}$$

and variance

$$\text{var}(Y^* \mid \beta, \sigma, G^*, D^*, F^*) = \sigma^2 \tag{6.7}$$

Furthermore, in Section 5 we obtained draws $(\beta^{(i)}, \sigma^{(i)})$, $i = 1, 2, ..., M$ (with $M = 45,000$) from the posterior pdf $f(\beta, \sigma \mid Y, X, R)$. Because a marginal pdf can be estimated as an average of conditional pdf's, it follows that an estimate of the predictive pdf is given by

$$\begin{aligned} \hat{f}(Y^* \mid Y, X, G^*, D^*, F^*, R) &= \frac{1}{M}\sum_{i=1}^{M} f(Y^* \mid \beta^{(i)}, \sigma^{(i)}, G^*, D^*, F^*) \\ &= \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{M} \frac{1}{\sigma^{(i)}} \exp\left\{-\frac{1}{2\sigma^{2(i)}}\left(Y^* - z\beta^{(i)}\right)^2\right\} \end{aligned} \tag{6.8}$$

A graph of this pdf can be created by evaluating (6.8) over a suitable grid of values for Y^*.

An estimate of its mean is given by

$$\hat{E}(Y^* \mid Y, X, G^*, D^*, F^*, R) = \frac{1}{M}\sum_{i=1}^{M} z\beta^{(i)} = z\overline{\beta} \tag{6.9}$$

Noting that an unconditional variance is equal to the average of the conditional variances plus the variance of the conditional means, an estimate of the variance is given by

$$\hat{\text{var}}(Y^* \mid Y, X, G^*, D^*, F^*, R) = \frac{1}{M}\sum_{i=1}^{M} \sigma^{2(i)} + \frac{1}{M-1}\sum_{i=1}^{M} (z\beta^{(i)} - z\overline{\beta})^2 \tag{6.10}$$

Other summary quantities can be obtained. For example, the probability that Y^* lies in a given interval can be obtained by averaging the conditional probabilities that Y^* lies in that interval. To obtain a predictive interval with, say, 95% probability content, one can draw values Y^* from the normal distribution $f(Y^* \mid \beta^{(i)}, \sigma^{(i)}, G^*, D^*, F^*)$, for $i = 1, 2, ..., M$; the 0.025 and 0.975 empirical quantiles from these draws give an estimate of the required interval.

6.3 With unknown rainfalls and no inequality restrictions

Turning now to the case where rainfall in one or more of the 1997 rainfall periods is unobserved, we write the joint pdf in equation (6.1) as

$$f(Y^*, \beta, \sigma, x^*, \mu, \tau \mid Y, X, x) = f(Y^*, \beta, \sigma \mid x^*, x, Y, X)\ f(x^* \mid \mu, \tau)\ f(\mu, \tau \mid X) \tag{6.11}$$

Given a noninformative prior without inequality constraints, it is possible to analytically integrate (β, σ) out of this expression so that the predictive pdf for yield can be written as

$$f(Y^* \mid Y, X, x) = \int\int\int f(Y^* \mid x^*, x, Y, X)\ f(x^* \mid \mu, \tau)\ f(\mu, \tau \mid X)\, dx^*\, d\mu\, d\tau \tag{6.12}$$

where the pdf $f(Y^* \mid x^*, x, Y, X)$ is identical to the t-distribution given in equation (6.3) for the observed rainfall case. Thus, to estimate $f(Y^* \mid Y, X, x)$, we can proceed as follows:

1. Draw $(\mu^{(i)}, \tau^{(i)}), i = 1, 2, \ldots, M$ from $f(\mu, \tau \mid X)$.

2. Draw $x^{*(i)}, i = 1, 2, \ldots, M$ from $f(x^* \mid \mu^{(i)}, \tau^{(i)})$.

3. For a grid of points for Y^*, average the *t*-distribution $f(Y^* \mid x^{*(i)}, x, Y, X)$ over the draws $x^{*(i)}$.

The methodology and results for steps 1 and 2 were given in Section 4. Step 3 can be written as

$$\hat{f}(Y^* \mid Y, X, x) = k \sum_{i=1}^{M} \left(\frac{1}{\sqrt{1 + z^{(i)}(Z'Z)^{-1} z^{(i)}{}'}} \right) \left[1 + \frac{(Y^* - z^{(i)}b)'(Y^* - z^{(i)}b)}{vs^2(1 + z^{(i)}(Z'Z)^{-1} z^{(i)}{}')} \right]^{-(v+1)/2}$$

where

$$k = \frac{\Gamma[(v+1)/2]}{\Gamma(v/2)(\pi vs^2)^{1/2}} \tag{6.13}$$

and $z^{(i)}$ is a (1×10) vector containing the constant and trend terms, the observed 1997 rainfalls and their squares, and the draws of the unobserved rainfalls $x^{*(i)}$ and their squares. Estimates of the posterior mean and variance are given by

$$\hat{E}(Y^* \mid Y, X, x) = \frac{1}{M} \sum_{i=1}^{M} z^{(i)} b = \bar{z} b \tag{6.14}$$

$$\hat{\text{var}}(Y^* \mid Y, X, x) = \left(\frac{vs^2}{v-2} \right) \frac{1}{M} \sum_{i=1}^{M} \left(1 + z^{(i)}(Z'Z)^{-1} z^{(i)}{}' \right) + \frac{1}{M-1} \sum_{i=1}^{M} (z^{(i)}b - \bar{z}b)^2 \tag{6.15}$$

Other summary quantities such as the probability that Y^* lies in a given interval, or a predictive interval with probability content of 0.95, can be obtained along the lines of the discussion below equations (6.8) to (6.10).

6.4 With unknown rainfalls and inequality restrictions

The final case that we consider is where one or more rainfalls are unobserved and inequality restrictions are imposed on the regression coefficients. We proceed as we did for the earlier inequality-restricted case (where rainfall was observed), except that we now average over the draws for both (β,σ) and x^*. The integral being estimated (see equation (6.1)) can be rewritten as

$$f(Y^* \mid Y, X, x, R) = \int\int\int\int\int f(Y^* \mid \beta, \sigma, x^*, x)\; f(\beta, \sigma \mid Y, X, R)\; f(x^* \mid \mu, \tau)\; f(\mu, \tau \mid X)\; d\beta\, d\sigma\, dx^*\, d\mu\, d\tau \qquad (6.16)$$

Again, redundant conditioning variables have been omitted. Equation (6.16) clarifies the sequential way in which draws are made from the various pdf's and the integral is estimated. Starting from the far right of the equation, (μ,τ) are drawn from $f(\mu,\tau \mid x)$; given these draws, x^* is drawn from $f(x^* \mid \mu,\tau)$. Draws (β,σ) are made from $f(\beta,\sigma \mid Y, X, R)$. Then, the normal pdf $f(Y^* \mid \beta,\sigma,x^*,x,Y,X)$ is averaged over the draws $(\beta^{(i)},\sigma^{(i)},x^{*(i)})$, $i = 1,2,...,M$. Specifically,

$$\hat{f}(Y^* \mid Y, X, x, R) = \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{M} \frac{1}{\sigma^{(i)}} \exp\left\{-\frac{1}{2\sigma^{2(i)}}\left(Y^* - z^{(i)}\beta^{(i)}\right)^2\right\} \qquad (6.17)$$

The estimated mean of the predictive pdf is

$$\hat{E}(Y^* \mid Y, X, x, R) = \frac{1}{M}\sum_{i=1}^{M} z^{(i)}\beta^{(i)} = \overline{z\beta} \qquad (6.18)$$

The estimated variance is

$$\hat{\text{var}}(Y^* \mid Y, X, x, R) = \frac{1}{M}\sum_{i=1}^{M} \sigma^{2(i)} + \frac{1}{M-1}\sum_{i=1}^{M} (z^{(i)}\beta^{(i)} - \overline{z\beta})^2 \qquad (6.19)$$

6.5 Results

The means and standard deviations of the predictive pdf's for 1997 yield for each shire and for each forecast time are given in Table 5. Realised yields and rainfalls are in the same table. Yields are in units of tonnes per hectare. Rainfall is measured relative to the sample average. We can make the following observations:

1. Predicted yield is highest for Irwin, followed by Northampton, Greenough, Chapman Valley and Mullewa. The average May-October rainfalls for these shires, in millimetres, are 385, 400, 395, 377 and 251, respectively. Thus, with the exception of Irwin, a ranking of shires on the basis of predicted yield corresponds with a ranking according to average growing-season rainfall.

2. Realised yields in 1997 are all greater than their corresponding predictive means with the exception of Irwin where the realisation is less than predicted. All realisations are within the effective ranges of their corresponding predictive pdf's. This fact can be confirmed for Mullewa and Irwin by placing their yields of 1.799 and 1.934, respectively, on Figures 3 and 4. Further discussion about these figures appears below.

3. In the remaining points we make observations on how the predictive means and standard deviations change as rainfall uncertainty is reduced, and relate some of these changes to whether realised rainfalls are above or below average. The effect of the inequality constraints is also mentioned. We conclude that the predictive pdf is an effective device for capturing regressor uncertainty introduced via the timing of the forecast, and prior information on regression coefficients. However, it is worth making the general observation that, for this data set, the timing of the forecast and the presence or otherwise of the inequality restrictions do not have relatively large effects on the predictive pdf's. Overall, rainfall is reliable; variations attributable to rainfall uncertainty are much less than variations attributable to the random error in the regression model. Similar remarks can be made about the inequality restrictions. They do correct some perverse directions of change, but their impacts are not of substantial magnitude relative to uncertainty in the error term.

4. To examine the effect of timing of the forecast on the mean of the predictive pdf, consider, for example, the inequality prior column for Mullewa. Before any rainfalls have been observed, the predictive mean is 1.629. After observing a below average germination rainfall of 0.578, the predictive mean falls to 1.530. Following an above average development-period rainfall of 1.329, it increases to 1.584. Then, an above average flowering rainfall of 1.038 increases it further to 1.650.

5. As expected, in all cases, extra rainfall uncertainty increases the standard deviation of the predictive pdf. This is illustrated in Figure 3 where the inequality-restricted pdf's are graphed for Mullewa. Note that the pdf's become less flat as more rainfalls are realised. Also, in line with point 4, the mean first shifts to the left and then back to the right.

6. The general effect of the inequality restrictions has been to reduce the predictive standard deviations in Northampton, Chapman Valley and Mullewa, and to increase the predictive standard deviations in Greenough and Irwin. The larger standard deviations in Greenough and Irwin can be attributed to the larger posterior means for σ that occur in the inequality-restricted case. Given that error uncertainty is large relative to coefficient uncertainty, it is relatively easy for an increase in σ to outweigh any improvements in the precision of estimation of the coefficients.
7. The inequality restrictions can make the relationships between predictive pdf's more realistic. Consider the noninformative prior results for Irwin shire. After observing a below average germination rainfall of 0.665, the predictive mean increases from 2.262 to 2.304. This unexpected outcome is corrected in the inequality-restricted case where the predictive mean decreases from 2.225 to 2.218. A comparison of the two Irwin pdf's, with and without inequality restrictions, and after observing germination rainfall, appears in Figure 4.

7. Concluding Remarks

Because rainfall at different times during the growing season is a major determinant of wheat yield, and yield forecasts need to be made before one or more rainfall periods have been realised, the forecasting problem can be viewed as a regression forecasting problem with uncertain regressors. By conditioning on observables, and averaging over unobservables, Bayesian inference provides a consistent and unified approach to this forecasting problem. Changes in the level of uncertainty that occur naturally through the timing of the forecast can be explicitly introduced into the predictive probability density function. This methodology has been developed and illustrated using shire level data for 5 shires in Western Australia.

Appendix: Metropolis-Hastings Algorithm (Random Walk Version)

Let θ be a vector of unknown parameters.

(1) Set θ equal to some starting value $\theta^{(0)}$ and compute the corresponding posterior pdf value $f\left(\theta^{(0)}\,|\,\text{data}\right)$, apart from the unknown normalizing constant. Any initial value, θ_0, in the feasible set can be used but it is usual to set $\theta^{(0)}$ equal to

the maximum likelihood estimate in order to make the algorithm more efficient in terms of the number of draws necessary to produce a reasonable characterisation of the posterior pdf.

(2) Commencing with $n = 0$, compute a potential (candidate) value

$$\theta^* = \theta^{(n)} + v$$

for draw $n\{n = 0,1,...\}$ where $v \sim N(0, kV)$. V is chosen as the maximum likelihood covariance matrix, but other covariance matrices could be used. The parameter k is chosen by trial and error to give an acceptance rate (defined below) of about 50%.

(3) Check whether θ^* lies in the feasible region. If it does not, $\theta^{(n)}$ becomes a draw $\left(\theta^{(n+1)} = \theta^{(n)}\right)$; increment n and return to step 2. Otherwise, proceed to step 4.

(4) Compute $f\left(\theta^* \,|\, \text{data}\right)$ and

$$r = f(\theta^* \,|\, \text{data}) \big/ f(\theta^{(n)} \,|\, \text{data})$$

the ratio for the candidate draw to that for the previous draw. If $r > 1$, the new draw is closer to the mode of the distribution and, conversely, $r < 1$ implies the new draw is further in the tail of the posterior pdf.

(5) If $r > 1$, θ^* becomes a draw $\left(\theta^{(n+1)} = \theta^*\right)$, increment n and return to step 2. If $r < 1$, proceed to step 6.

(6) Generate a uniform random number, say ε, from $(0,1)$. If $\varepsilon \le r < 1$, θ^* becomes a draw; set $\theta^{(n+1)} = \theta^*$ and return to step 2. If $r < \varepsilon$, $\theta^{(n)}$ becomes a draw $\left(\theta^{(n+1)} = \theta^{(n)}\right)$; return to step 2. Thus, if $r \ge 1$, the new value is accepted as a draw. When $r < 1$, the algorithm does not necessarily move towards the tail of the distribution. Thus, more draws occur in regions of high probability, and fewer draws occur in regions of low probability.

Repeat the algorithm, say 50,000 times. The acceptance rate is the proportion of times θ^* is accepted as a draw. Markov Chain Monte Carlo theory guarantees that, after a large number of draws, say 5,000, the remaining draws will be from the posterior pdf $f(\theta \mid \text{data})$. These draws are not independent, but, from the law of large numbers, can nevertheless be used to consistently estimate to required pdfs and their moments. For further details, see for example, Geweke (1999).

References

Chatfield, C. (1993), "Calculating Interval Forecasts", *Journal of Business & Economic Statistics*, 11, 121-135.

Coelli, T.J (1992), "Forecasting Wheat Production Using Shire Level Data",paper contributed to the 36th Annual Conference of the Australian Agricultural Economics Society, Canberra, Feb. 10-12.

Feldstein, M.S. (1971), "The Error of Forecast in Econometric Models when the Forecast-period Exogenous Variables are Stochastic", *Econometrica*, 39, 55-60.

Geweke, J. (1999), "Using Simulation Models for Bayesian Econometric Models: Inference, Development and Communication", *Econometric Reviews,* 18, 1-74.

Judge, G.G., R.C. Hill, W.E. Griffiths, H. Lütkepohl and T. Lee (1988), *Introduction to the Theory and Practice of Econometrics*, second edition, New York: John Wiley and Sons.

McCullough, B.D. (1996), "Consistent Forecast Intervals when the Forecast-period Exogenous Variables are Stochastic", *Journal of Forecasting*, 15, 293-304.

Newton, L.S. (2001), *Bayesian Prediction: Alternative Models for Forecasting Shire Level Wheat Yield*, Ph.D. thesis in preparation, The University of New England.

Stephens, D.J. (1995), *Crop Yield Forecasting over Large Areas In Australia*, Ph.D. thesis, Murdoch University.

Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, New York: John Wiley and Sons.

Table 1: Predictive and Posterior Means and (Standard Deviations) for Rainfall and its Parameters

	Shire				
	Northampton	Chapman Valley	Mullewa	Greenough	Irwin
Germination					
G^*	1.000 (0.385)	0.998 (0.365)	1.000 (0.394)	1.000 (0.397)	1.000 (0.370)
μ	0.993 (0.059)	0.995 (0.055)	0.990 (0.062)	0.990 (0.062)	0.995 (0.056)
τ	0.387 (0.046)	0.366 (0.043)	0.399 (0.050)	0.400 (0.050)	0.373 (0.043)
Development					
D^*	1.002 (0.377)	1.005 (0.383)	0.997 (0.425)	1.003 (0.328)	1.001 (0.354)
μ	0.993 (0.057)	0.992 (0.059)	0.983 (0.068)	0.997 (0.048)	0.997 (0.053)
τ	0.378 (0.044)	0.387 (0.047)	0.438 (0.056)	0.323 (0.036)	0.354 (0.042)
Flowering					
F^*	1.002 (0.461)	1.000 (0.429)	0.997 (0.669)	1.002 (0.412)	0.999 (0.420)
μ	0.966 (0.082)	0.980 (0.072)	0.490 (0.567)	0.985 (0.067)	0.983 (0.068)
τ	0.487 (0.069)	0.445 (0.058)	0.930 (0.271)	0.420 (0.054)	0.431 (0.055)

Table 2: Posterior Means and (Standard Deviations) for Regression Parameters with a Noninformative Prior

Variable	Northampton	Chapman Valley	Mullewa	Greenough	Irwin
constant	0.144 (0.315)	0.089 (0.319)	-0.313 (0.291)	0.404 (0.274)	0.968 (0.454)
T	36.078 (22.068)	22.274 (22.172)	22.709 (22.404)	32.420 (18.968)	37.907 (24.855)
$T^{(2)}$	-1.951 (1.066)	-1.372 (1.058)	-1.538 (1.092)	-1.557 (0.916)	-2.585 (1.186)
$T^{(3)}$	0.038 (0.015)	0.028 (0.014)	0.030 (0.015)	0.029 (0.012)	0.050 (0.016)
G^{*}	0.660 (0.290)	0.889 (0.339)	1.007 (0.464)	0.299 (0.259)	0.244 (0.412)
G^{*2}	-0.341 (0.113)	-0.435 (0.147)	-0.397 (0.205)	-0.201 (0.103)	-0.178 (0.171)
D^{*}	0.135 (0.330)	0.217 (0.347)	0.309 (0.280)	0.391 (0.350)	-0.329 (0.543)
D^{*2}	-0.075 (0.134)	-0.072 (0.147)	-0.091 (0.114)	-0.182 (0.153)	0.088 (0.225)
F^{*}	0.496 (0.307)	0.201 (0.338)	0.511 (0.160)	-0.004 (0.309)	0.137 (0.349)
F^{*2}	-0.187 (0.148)	-0.015 (0.159)	-0.131 (0.052)	0.075 (0.141)	-0.030 (0.169)
σ	0.181 (0.022)	0.168 (0.020)	0.185 (0.022)	0.159 (0.019)	0.205 (0.025)

Table 3: Posterior Means and (Standard Deviations) for Regression Parameters with an Inequality-restricted Prior

Variable	Northampton	Chapman Valley	Mullewa	Greenough	Irwin
constant	-0.152 (0.272)	-0.024 (0.305)	-0.326 (0.290)	0.141 (0.245)	0.085 (0.379)
T	40.857 (21.116)	21.028 (22.093)	22.385 (22.592)	29.857 (18.990)	40.439 (25.855)
$T^{(2)}$	-2.114 (1.026)	-1.273 (1.050)	-1.521 (1.089)	-1.356 (0.905)	-2.535 (1.215)
$T^{(3)}$	0.039 (0.014)	0.026 (0.014)	0.030 (0.015)	0.026 (0.012)	0.048 (0.017)
G^*	0.897 (0.221)	1.014 (0.298)	1.011 (0.463)	0.647 (0.201)	0.768 (0.364)
G^{*2}	-0.417 (0.099)	-0.472 (0.140)	-0.399 (0.205)	-0.304 (0.094)	-0.358 (0.173)
D^*	0.308 (0.273)	0.249 (0.295)	0.343 (0.264)	0.452 (0.310)	0.440 (0.391)
D^{*2}	-0.120 (0.124)	-0.077 (0.133)	-0.101 (0.111)	-0.187 (0.145)	-0.192 (0.187)
F^*	0.481 (0.301)	0.208 (0.325)	0.495 (0.157)	-0.075 (0.302)	0.241 (0.377)
F^{*2}	-0.187 (0.147)	-0.021 (0.154)	-0.126 (0.051)	0.106 (0.140)	-0.067 (0.184)
σ	0.182 (0.022)	0.167 (0.020)	0.185 (0.022)	0.163 (0.020)	0.218 (0.026)

Table 4: Posterior Probabilities for the Inequality Regions Given a Noninformative Prior.

	Shire				
	Northampton	Chapman Valley	Mullewa	Greenough	Irwin
Germination	0.408	0.587	0.991	0.101	0.157
Development	0.439	0.795	0.924	0.619	0.118
Flowering	0.964	0.993	1.000	0.980	0.829

Table 5: Means and (Standard Deviations) from Predictive pdf's for Yield and Yield and Rainfall Realisations

	Northampton		Chapman Valley		Mullewa		Greenough		Irwin	
Unknown rainfalls	Noninf prior	Inequality prior	Noninf prior	Inequality prior	Noninf prior	Inequality prior	Noninf prior	Inequality prior	Noninf prior	Inequality prior
none	2.129	2.161	1.793	1.759	1.649	1.650	1.999	1.914	2.321	2.279
	(0.235)	(0.225)	(0.212)	(0.206)	(0.236)	(0.233)	(0.198)	(0.198)	(0.246)	(0.260)
F^*	2.084	2.116	1.759	1.726	1.580	1.584	1.977	1.898	2.295	2.241
	(0.249)	(0.240)	(0.229)	(0.222)	(0.282)	(0.279)	(0.211)	(0.212)	(0.254)	(0.272)
D^*, F^*	2.072	2.114	1.738	1.700	1.531	1.530	1.966	1.894	2.304	2.218
	(0.252)	(0.244)	(0.233)	(0.225)	(0.289)	(0.287)	(0.215)	(0.217)	(0.269)	(0.276)
G^*, D^*, F^*	2.120	2.125	1.775	1.761	1.630	1.629	1.935	1.947	2.262	2.225
	(0.257)	(0.258)	(0.246)	(0.243)	(0.311)	(0.308)	(0.225)	(0.227)	(0.281)	(0.287)
Realised yields										
	2.232		1.843		1.799		2.235		1.934	
Realised rainfalls (relative to sample average)										
F^*	1.051		1.181		1.038		1.213		1.296	
D^*	0.838		1.181		1.329		0.848		0.985	
G^*	1.504		0.554		0.578		0.484		0.665	

Figure 2: Estimated Trend for Greenough Shire

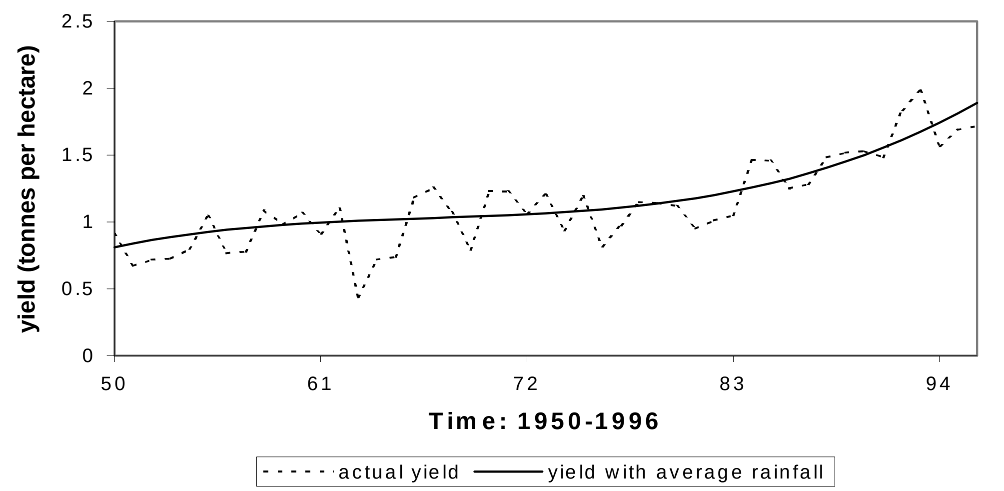
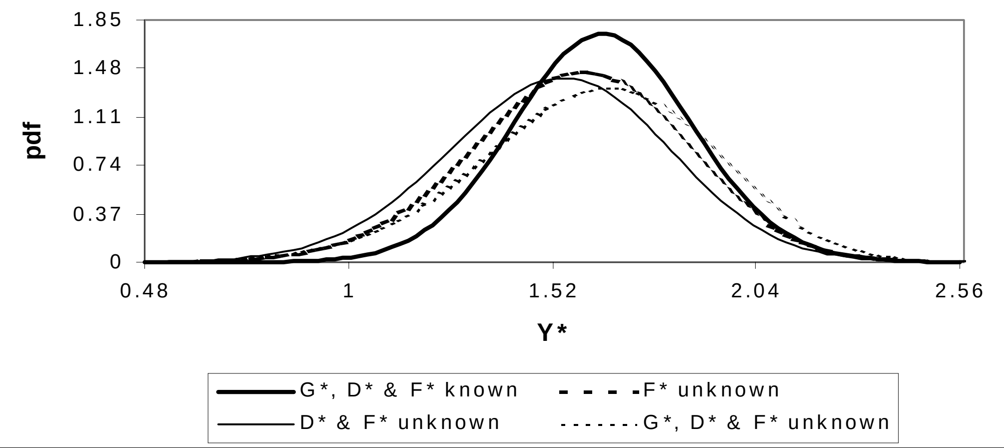


Figure 3: Inequality Restricted Predictive PDFs for Mullewa



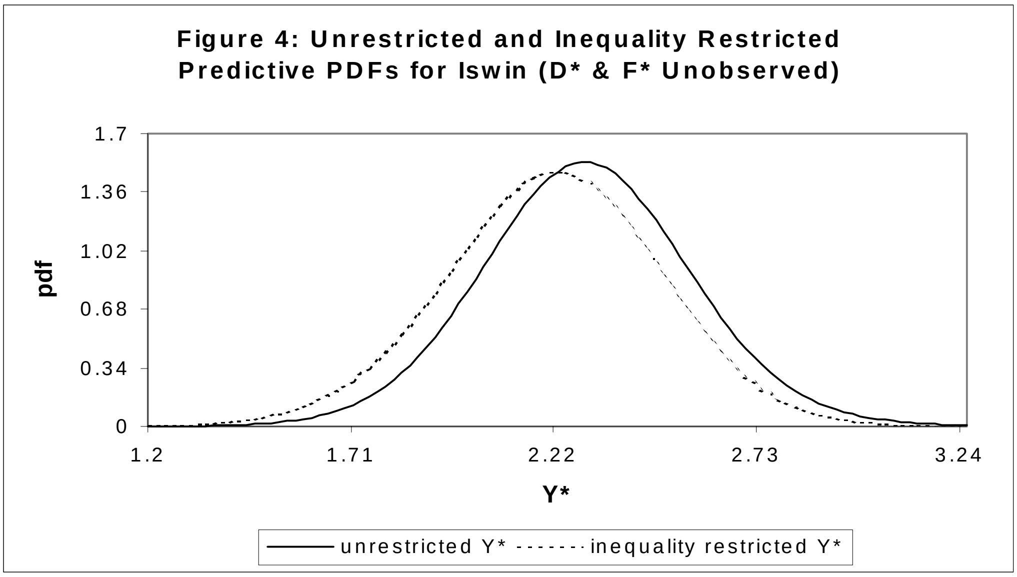


Figure 4: Unrestricted and Inequality Restricted
Predictive PDFs for lswin (D* & F* Unobserved)
pdf
1.7
1.36
1.02
0.68
0.34
0
1.2
1.71
2.22
2.73
3.24
Y*
unrestricted Y*
inequality restricted Y*