Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
International Journal of Forecasting 26 (2010) 397–412
international journal of forecasting
www.elsevier.com/locate/ijforecast

# Predictive densities for models with stochastic regressors and inequality constraints: Forecasting local-area wheat yield

William E. Griffiths[a,*], Lisa S. Newton[b], Christopher J. O'Donnell[c]

[a] *Department of Economics, University of Melbourne, Vic 3010, Australia*
[b] *Queensland Competition Authority, GPO Box 2257, Brisbane, Qld 4001, Australia*
[c] *School of Economics, University of Queensland, St Lucia, Qld 4072, Australia*

**Abstract**

Forecasts from regression models are frequently made conditional on a set of values for the regressor variables. We describe and illustrate how to obtain forecasts when some of those regressors are stochastic and their values have not yet been realized. The forecasting device is a Bayesian predictive density which accommodates variability from an unknown error term, uncertainty from unknown coefficients, and uncertainty from unknown stochastic regressors. We illustrate how the predictive density of a forecast changes as more regressors are observed and therefore fewer are unobserved. An example where the local-area wheat yield depends on the rainfall during three periods – germination, growing and flowering – is used to illustrate the methods. Both a noninformative prior and a prior with inequality restrictions on the regression coefficients are considered. The results show how the predictive density changes as more rainfall information becomes available.


*Keywords:* Bayesian forecasting; Inequality restrictions; Random regressors; Rainfall distributions; Truncated distributions

## 1. Introduction

The linear regression model is a common vehicle for forecasting future values of economic variables. In the traditional textbook treatment of linear-model forecasting (see, e.g., Wooldridge, 2009, p. 208), future values of a dependent variable are predicted for given values of the explanatory variables (regressors). Expressions for both point and interval forecasts are obtained, with the interval forecasts using a standard error that reflects both uncertainty about the future value of the error terms and the sampling error from estimating the coefficients. A major simplification implicit in this treatment is the assumption that future values of the regressors are known. Because the analysis conditions on these values, any uncertainty attached to them is ignored in the construction of interval forecasts. This limitation is not a serious one if the regressors are all policy variables whose

* Corresponding author.
*E-mail addresses:* wegrif@unimelb.edu.au (W.E. Griffiths), lisa.haselwood@qca.org.au (L.S. Newton), c.odonnell1@uq.edu.au (C.J. O'Donnell).

0169-2070/$ - see front matter © 2010 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2009.12.008

values are set by a decision maker. For example, a retail store investigating sales of its product might set different prices and use different advertising strategies. However, there are many circumstances where some if not all of the future values of the explanatory variables are not known with certainty. Retail store sales might also depend on the weather or the behaviour of competitors, neither of which can be predicted with certainty. When there is a lack of knowledge about the future values of stochastic regressors, forecasts should have forecast intervals that reflect this lack of knowledge: the intervals should be wider than traditional intervals that condition on known values of the regressors.

In the sampling-theory literature there is general recognition that independent forecasts of stochastic regressors need to be provided, but there does not seem to be a unified framework for incorporating the consequent uncertainty into forecast intervals. McCullough (1996) makes some progress in this direction: he suggests using the bootstrap to obtain consistent forecast intervals, overcoming an inconsistency described by Feldstein (1971). Various approaches to obtaining forecast intervals for a range of models have been reviewed by Chatfield (1993). Lam and Veall (2002) illustrate the inaccuracy of normal-distribution based forecast intervals when the errors are non-normal. Bayesian inference, in contrast to the different sampling-theory approaches, provides a single unified framework for obtaining point and interval forecasts that reflect all sources of uncertainty. The Bayesian tool that incorporates the different sources of uncertainty is the predictive density function. This density function is the probability density function (pdf) for future values of the variable of interest, conditional only on past observables. It is obtained by forming the joint density function for all unobservables, including the regression coefficients, the future values of stochastic regressors, and future values of the dependent variable, all conditional on past observations, and then integrating out all unknowns with the exception of the forecast variable. The marginal pdf that results is the predictive density function; integrating out the other unobservables, rather than conditioning on them or their estimates, means that the predictive pdf reflects the uncertainty associated with the unobservables. Once the predictive pdf has been obtained, information from it can be presented in a number of ways. For point forecasts it is common to assume implicitly that a quadratic loss function is appropriate, in which case the mean of the predictive density is taken as the optimal point forecast. The reliability of that forecast, or a general idea of the uncertainty of any forecast, can be obtained by presenting a graph of the pdf, a measure of its dispersion such as the standard deviation, or a forecast interval. The forecast interval has a similar interpretation to a sampling-theory forecast interval, but it accommodates all sources of uncertainty, and may be influenced by prior information placed on the parameters of the regression model. Moreover, inferences from the predictive density are finite sample inferences. Asymptotic approximations are not necessary, as is typically the case with sampling theory inferences in models of similar levels of complexity.

The objective of this paper is to describe and illustrate how to obtain a predictive pdf for a linear or non-linear regression model with stochastic regressors. Two types of prior information are considered. One is a noninformative prior where past sample information dominates both the resulting posterior density for the coefficients and the predictive density. The other prior is an inequality-restricted prior, where the coefficients are assumed to lie within a restricted region, but the prior is noninformative otherwise. The latter prior is an important one in economics, as economic theory often dictates sign restrictions, as well as more complex inequalities on regression coefficients; see for example Barnett and Serletis (2008) and Griffiths, O'Donnell and Tan-Cruz (2000).

As far as we are aware, a predictive density which incorporates both inequality information on the coefficients and uncertainty from stochastic regressors has not previously been considered in the literature. Zellner and Park (1987) derive the predictive pdf under the assumption that the dependent and independent variables follow a multivariate normal distribution. They also derive expressions for the moments of the predictive pdf when the normality assumption is relaxed. Knight, Sirmans, Gelfand and Ghosh (1998) show how to use the Gibbs sampler to generate observations from the posterior pdf for linear regression coefficients and for any missing sample observations on the regressors. They mention that their methodology can readily be extended to the case where the dependent variable is unobserved and the