Working Papers in Econometrics and Applied Statistics

FLEXIBLE DISTRIBUTED LAGS

Duangkamon Chotikapanich

and

William E Griffiths

No. 105 – July 1999

Working Papers in Econometrics and Applied Statistics
Department of Econometrics
University of New England
Armidale, NSW 2351, Australia.

http://www.une.edu.au/econometrics/emetwp.htm

ISSN 0157-0188
ISBN 1 86389 597 3

FLEXIBLE DISTRIBUTED LAGS*

Duangkamon Chotikapanich
Curtin University of Technology

William E Griffiths
University of New England

Abstract

In econometrics there is a long history of using continuous functions to force distributed lag coefficients to behave in an economically accepted way. For example, geometrically declining lags have often been used to model coefficients that we believe should be declining. Polynomial lags have been used to model lag coefficients expected to increase and then decrease. In this paper a more flexible way of imposing such prior information is investigated. Inequality constraints are used to impose knowledge about the relative magnitudes of coefficients without forcing them to lie on a smooth continuous curve. A Metropolis algorithm is used to get posterior density functions for the lag coefficients and functions of those coefficients for the Nerlove orange data and the Almon capital expenditures data.

June 22, 1999

*Paper presented at the George Judge Conference held at the University of Illinois, Champaign, Illinois on May 1-2, 1999. The authors thank Dale Poirier and other participants at the Conference for helpful suggestions, and Chris O'Donnell for useful discussions.

FLEXIBLE DISTRIBUTED LAGS

1. INTRODUCTION

Distributed lag models have played a prominent role in numerous applications in the economics and agricultural economics literature. Early examples include the studies by Nerlove (1956,1958) on the response of agricultural supply to price, the study on capital appropriations and expenditures by Almon (1965) and the response of capital investment to various aspects of the economic environment (Koyck 1954, Jorgenson 1965). Textbook treatments of distributed lag models appear in Judge et.al. (1988, Ch. 17), Greene (1997, Ch. 17), Davidson and MacKinnon (1993, Ch. 19) and Baltagi (1999, Ch. 6). The distributed-lag literature has been surveyed by Griliches (1967), Dhrymes (1971), Hendry, Pagan and Sargan (1984) and Judge et.al. (1985). Some recent results have been derived by Dufour and Kiviet (1998).

One of the main features of many distributed-lag applications is the desire to have lag weights that are declining. It is often thought that the greatest impact of the explanatory variable x_t on a dependent variable y_t is the immediate one with the impact declining for values of the explanatory variable further in the past. That is, $\partial y_t / \partial x_{t-i}$ declines as i increases. In some cases (for example, area crop response where a decision about area sown is made prior to observation of current price) the first and largest impact is assumed to be for i = 1, but the presumed declining nature of the distributed lag is still the same. In other applications, it is recognized that the lag weights are likely to increase to a maximum and then decline. For example, in the relationship between capital expenditures and capital appropriations, the major effect of an appropriations decision on expenditure is not likely to be felt immediately. Most investment projects require a start-up time when the expenditures from an appropriation are likely to be increasing; then, as the work on the investment projects draws to a close, the expenditures from the appropriation are likely to decline (Almon, 1965).

To capture lag weights that continually decline, or that increase up to a point in time and then decline, distributed lag models have forced the lag weights to lie on an appropriately-chosen continuous function. Declining lag weights are frequently captured using a geometrically declining lag, or a polynomial lag of degree one, with much of the popularity of the geometrically-declining lag being attributable to its justification in terms of partial-adjustment or adaptive-expenditures modelling (Judge et. al., 1988, p.735). Lag weights which increase and then decrease can be captured with polynomial lags of degrees 2 or 3; or, the added restriction that the weights be nonnegative can be handled via an exponential lag (Lutkepohl, 1981). Other lag schemes, documented in the surveys referred to above, have also been suggested.

A frequent consequence of not forcing the lag weights to lie on a continuous function, and using unrestricted estimation instead, is an estimated lag pattern that is not acceptable relative to our prior views. High correlations between lagged values of an explanatory variable can lead to least-squares lag-weight estimates with high variances. High variances can, in turn, produce lag patterns that do not conform to reasonable prior expectations. It is the existence of high correlations between lagged values of an explanatory variable, and the fact that we have strong prior information about acceptable lag patterns, that have motivated the use of continuous functions for modelling lag structures. Imposing such structures on the estimated parameters of a model makes the estimates (or the relationship between them) conform more closely to our a priori expectations. It is also hoped that the restrictions are true, or at least not "too false" (in the context employed by Toro-Vizcorrondo and Wallace, 1968), such that the distributed lag estimates have smaller mean-squared errors than their corresponding unrestricted counterparts.

The purpose of this paper is to introduce distributed lag models which place less structure on the weights than any of the alternatives considered so far in the literature, but which still preserve the essential features of typical prior information. It is our belief that the essential a priori restrictions that many applied researchers wish to impose can be written in terms of inequality restrictions. When declining lag weights are assumed, researchers usually do not feel strongly about the rate of decline or whether the decline over several periods has to be smooth. Nevertheless, they proceed with structures that ensure a smooth decline and which are more restrictive than necessary because of the convenience of tools that impose those structures. Similar comments can be made for lags that increase, then decline. The tool which we are suggesting is one that insists the parameter estimates satisfy inequality restrictions, but nothing more. As we will see in the examples that follow, both declining lag-weights, and weights that increase and then decline, can be readily expressed in terms of inequality restrictions. Furthermore, estimation of the lag-weights subject to inequality restrictions is best achieved within the framework of Bayesian inference. Much of the early work on inequality–restricted estimation was carried out within the sampling-theory framework by George Judge and his colleagues. Judge and Takayama (1966) showed how quadratic programming can be utilized to obtain least-squares estimates subject to inequality restrictions. Studies which evaluate the sampling-theory properties of such estimators for a number of special cases are reviewed in Judge and Yancey (1986). However, as noted by Geweke (1986), sampling-theory estimation subject to inequality restrictions has two undesirable characteristics. It produces boundary solutions whose estimates can be uninteresting and not very informative, and our ability to carry out finite-sample inference is restricted to a few special cases. On the other hand, the posterior means from Bayesian inference provide a set of estimates which are not boundary solutions, and the posterior density functions of the lag weights form a basis for finite-sample inference.

Geweke (1986) described how truncated posterior density functions that are a consequence of inequality restrictions can be estimated via importance sampling. The importance-sampling algorithm is equivalent to drawing values of the parameter vector from a (nontruncated) multivariate *t*-distribution, and discarding draws which do not satisfy the restrictions. One problem with this approach, when there are several inequality restrictions, is that the probability of drawing a feasible vector can be almost zero, making the total number of draws needed for reasonable accuracy prohibitive. A Gibbs-sampling algorithm which overcomes this problem for linear inequality restrictions has been suggested by Geweke (1991, 1996). This algorithm could be applied to our inequality-restricted distributed lag model. However, as an alternative, we suggest a Metropolis algorithm which we believe is easier to implement. We have used this algorithm before for imposing inequality restrictions in the context of nonlinear seemingly unrelated regressions (Griffiths and Chotikapanich, 1997). It has the added advantage of being able to handle nonlinear inequality constraints, although such constraints do not exist with the distributed lag models studies in this paper. The Gibbs-sampling and Metropolis algorithms come under the general heading of Markov Chain Monte Carlo which has turned many previously intractable Bayesian applications into a practical reality. For access to the literature, and an appreciation of the wide variety of applications that can be handle via Markov Chain Monte Carlo, see Tanner (1993), Albert and Chib (1996), Chib and Greenberg (1996), Gilks et.al. (1996), and Geweke (1999).

In the examples that follow we compare Bayesian and sampling-theory inequality restricted estimates. The sampling-theory estimates can be computed using quadratic programming (Judge and Takayama, 1966). However, given that many researchers may find it more convenient to use a standard econometric software package, we demonstrate how these estimates can be computed by applying nonlinear least squares to a transformed model. The general model and estimation details are described in Section 2. A distributed lag model with declining lag weights is considered in Section 3. In Section 4 a model with weights that increase and then decline is considered.

2. THE MODEL AND ESTIMATION

A general finite-lag distributed lag model can be written as:

$$y_t = \alpha_0 x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_n x_{t-n} + z'_t \beta + e_t \qquad t = 1, 2, \ldots, N \tag{1}$$

where e_t is independently and identically distributed as $N(0, \sigma^2)$;

$\alpha_0, \alpha_1, \cdots, \alpha_n$ are lag weights where n is known or set sufficiently large;

z_t is a $(p \times 1)$ vector containing other explanatory variables, one of which is usually a constant;

β is a $(p \times 1)$ vector containing other parameters.

Equation (1) can be written more compactly as:

$$\begin{aligned} y_t &= x_t'\alpha + z_t'\beta + e_t \\ &= \begin{pmatrix} x_t' & z_t' \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + e_t \\ &= w_t'\theta + e_t \end{aligned}$$

where w_t and θ are of order $(K \times 1)$ and $K = p + n + 1$.

The density function for the data generation process, which is identical to the likelihood function for the unknown parameters θ and σ is given by

$$f(y \mid \theta, \sigma) = \frac{1}{(2\pi)^{-T/2}} \frac{1}{\sigma^{-T}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{t=n+1}^{N} (y_t - w_t'\theta)^2 \right\} \tag{2}$$

Note that because of the lags of the explanatory variables, *n* observations are "lost" in creating $x_{t-1}, x_{t-2}, \ldots, x_{t-n}$, and therefore $T = N - n$.

The necessary steps to proceed with Bayesian estimation are (1) specification of a prior probability density function (pdf) that accommodates the inequality restrictions on the lag-weights, (2) derivation of the joint posterior pdf for all the unknown parameters through application of Bayes' Theorem, and (3) isolation of information from the joint posterior pdf (marginal posterior pdfs, posterior means and standard deviations) so that meaningful results can be presented.

For specification of a prior pdf, a uniform distribution is assigned to all parameters in θ, with the parameters in α constrained to lie within the feasible region defined by the inequality restrictions. The uniform inequality restricted prior implies that values of the parameters that *do not* satisfy the inequalities are impossible, and that all values that *do* satisfy the inequalities are *equally likely*. Examples of a feasible region, that we denote by *R*, are given in the next two sections. The expression for the prior distribution is

$$f(\theta) \propto \text{constant} \times I(R)$$

where $I(R)$ is an indicator function with the following properties.

$$I(R) = \begin{cases} 1 & \text{if } \theta \text{ falls within region } R \\ 0 & \text{if } \theta \text{ falls outside region } R \end{cases}$$

Assuming *a priori* independence of θ and σ, and using the conventional noninformative prior $f(\sigma) \propto \sigma^{-1}$ (Judge et al, 1988, p.150), the joint posterior pdf for θ and σ can be written as

$$f(\theta,\sigma) = f(\theta)\, f(\sigma) \propto \frac{I(R)}{\sigma} \tag{3}$$

Using Bayes' Theorem to combine the prior pdf in equation (3) and the likelihood function in equation (2) yields the joint posterior pdf

$$f(\theta,\sigma \mid y) \propto \frac{I(R)}{\sigma^{T+1}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=n+1}^{N} (y_t - w_t'\theta)^2\right\} \tag{4}$$

Given equation (4) is the joint posterior pdf for all the unknown parameters, in a Bayesian investigation it is the source of joint inferences about these parameters. Our interest centers on θ. For this case the marginal posterior pdf for θ is the relevant posterior pdf. To obtain it σ is integrated out from equation (4) to yield

$$f(\theta \mid y) = \int f(\theta,\sigma \mid y)\, d\sigma \propto \left[\sum_{t=n+1}^{N} (y_t - w_t'\theta)^2\right]^{-T/2} I(R) \tag{5}$$

Equation (5) contains information on a number of quantities of likely interest. Marginal posterior distributions for each of the lag weights, for the other parameters in the vector θ, and for functions of the lag weights, such as their sum, are likely to be useful. Also, posterior means and standard deviations are a convenient way of summarizing our information about such quantities. In the examples that follow, it is the posterior means of the lag weights which we take as point estimates to compare with estimates obtained using other techniques. All these quantities of interest are defined as integrals involving equation (5). However, these integrals are analytically intractable because the endpoints defined by the region *R* depend on θ. To overcome the analytical intractability, we estimate the various quantities using draws from the pdf in equation (5). Marginal posterior pdfs for the lag weights (for example) can be estimated by putting the draws in histograms. Estimates of the posterior means and standard deviations are given by the sample means and standard deviations of the draws. To obtain the draws we use what is called a random-walk Metropolis-Hasting algorithm. References given in the Introduction provide access to the Markov Chain Monte Carlo literature which has more details on this algorithm. The steps that we employ are as follows.

The Metropolis-Hastings Algorithm

1. Select feasible initial values for the elements of θ, say θ_0. One possibility is the least-squares estimates from an appropriately chosen polynomial lag. Perform the remaining steps with *i* set equal to 0.
2. Compute a value for the kernel of $\log f(\theta_i \mid y)$ given in equation (5).
3. Generate *d* from $N(0, kV)$. The choice of *V* and *k* are discussed below.
4. Compute a candidate draw $\theta^* = \theta_i + d$.

5. If $\boldsymbol{\theta}^*$ falls outside the feasible region, the next draw is identical to the previous one. That is, set $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i$, set $i = i + 1$ and return to step 2; otherwise, proceed with step 6.

6. Compute a value for the kernel of $\log f(\boldsymbol{\theta}^* \mid y)$ and the ratio of the pdfs

$$r = \frac{f(\boldsymbol{\theta}^* \mid y)}{f(\boldsymbol{\theta}_i \mid y)} = \exp\left\{\log f(\boldsymbol{\theta}^* \mid y) - \log f(\boldsymbol{\theta}_i \mid y)\right\}$$

7. If $r \geq 1$, the next draw becomes the candidate draw. That is, set $\theta_{i+1} = \theta^*$, set $i = i + 1$ and return to step 2; otherwise proceed with step 8.

8. If $r \leq 1$, the next draw is $\boldsymbol{\theta}^*$ with probability *r* and θ_i with probability 1-*r*. That is, generate a uniform random variable, say ν from the interval (0, 1). If $\nu \leq r$, set $\theta_{i+1} = \theta^*$. If $\nu > r$, set $\theta_{i+1} = \theta_i$. Set $i = i + 1$ and return to step 2.

The Metropolis-Hastings algorithm provides a means for moving around the parameter space and drawing observations consistent with $f(\theta \mid y)$. The vector *d* in step 3 represents a potential change from the last drawing of θ; the potential new value θ^* is given by the random walk process in step 4. The choice of covariance matrix *V* in step 3 is not critical. We chose the one from unrestricted least-squares estimation. In line with convention, the value *k* used for step 3 was set so that the acceptance rate for θ^* was approximately 0.5. The first check to see whether a potential new draw θ^* is acceptable is that in step 5; if any of the elements in θ^* do not satisfy the inequality constraints, the whole vector is rejected. In steps 7 and 8 we accept it with probability given by the ratio of the two densities. Thus, the procedure explores the posterior pdf yielding a relatively high proportion of observations in regions of high probability and a relatively low proportion of observations in regions of low probability.

Because Markov Chain Monte Carlo procedures produce observations that are correlated, their sample means and variances are not as efficient as they would be from independent observations; larger samples are needed to achieve a desired level of accuracy. Also, the draws are not genuine draws from $f(\theta \mid y)$ until the Markov Chain has converged. For this reason, a number of the early draws, referred to as a "burn-in" are discarded. In the two studies in this paper, 20,000 draws were discarded for the burn-in, 160,000 subsequent draws were used for the declining lag imposed on the orange data, and 100,000 subsequent draws were used for the lag imposed on the Almon data.

Restricted Maximum Likelihood Estimation

To obtain maximum likelihood estimates subject to the inequality restrictions, we apply nonlinear least squares to a transformed model. The model is reparameterized so that the inequality restrictions on the original parameters can be written in terms of nonnegative restrictions on single parameters in the transformed model. Nonnegativity of the transformed parameters is ensured by using nonlinear least squares with squared parameters. Let D be a $(k \times k)$ non-singular matrix such that the inequality restrictions can be written as

$$D\theta \geq C \tag{6}$$

where the elements of C are $-\infty$ for parameters not subject to inequality restrictions and 0 for linear functions of parameters that are subject to inequality restrictions. Examples of the matrix D will be given in the applications which follow. Let $\gamma = D\theta$ and rewrite the model $y_t = w_t'\theta + e_t$ as

$$\begin{aligned} y_t &= w_t' D^{-1} D \theta + e_t \\ &= q_t' \gamma + e_t \end{aligned} \tag{7}$$

where $q_t' = w_t' D^{-1}$. For the elements of γ that are restricted to be positive (the corresponding elements of C are zero), define $\gamma_k = \delta_k^2$. Estimate equation (7) via nonlinear least squares with the relevant γ_k replaced by δ_k^2. This strategy yields estimates $\hat{\delta}_k$ where restrictions are involved, and estimates $\hat{\gamma}_j$ where they are not. Find $\hat{\gamma}_k = \hat{\delta}_k^2$. Maximum likelihood estimates subject to the inequality restrictions are then given by

$$\hat{\theta} = D^{-1} \hat{\gamma} \tag{8}$$

3. A DECLINING LAG

For our first illustration we reconsider a well known and often cited study by Nerlove and Waugh (1961) who investigated the effect of advertising expenditure on the demand for oranges over the period 1910-1959. Their study was considered further by Berndt (1991). They related q_t (per capita deliveries of oranges in boxes in year *t*) to p_t (price of oranges per box), y_t (per capita disposable income of consumers), and a_t (per capita advertising expenditures of oranges by Sunkist Growers and the Florida Citrus Commission). The basic model specification that was used in this study is:

$$q_t = \beta_0 + \beta_1 p_t + \beta_2 y_t + \alpha_0 a_t + \alpha_1 a_{t-1} + \alpha_2 a_{t-2} + \cdots + \alpha_8 a_{t-8} + e_t$$

The lag length of 8 was chosen after experimentation with different lag lengths for both unrestricted and linearly declining lags. If we assume that the greatest effect of an advertising expenditure is felt immediately, and the effect gradually declines as we move further into the future, and never becomes negative, then the feasible region of the parameter space is

$$R = \{\theta \mid \alpha_0 \geq \alpha_1 \geq \alpha_2 \geq \alpha_3 \geq \alpha_4 \geq \alpha_5 \geq \alpha_6 \geq \alpha_7 \geq \alpha_8 \geq 0\}$$

Equation (6) can be written as

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \\ \alpha_7 \\ \alpha_8 \end{bmatrix} \geq \begin{bmatrix} -\infty \\ -\infty \\ -\infty \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Table 1 contains four sets of estimates:

1. Unrestricted least-squares estimates (OLS).
2. Least-squares estimates where the lag weights are constrained to lie on a polynomial of degree one (equality restricted LS).
3. The maximum likelihood estimates subject to inequality restrictions (inequality restricted LS).
4. The posterior means of the parameters (Bayesian).

The values in parentheses are standard errors for the sampling-theory estimates and posterior standard deviations for the Bayesian estimates. No standard errors are given for the inequality restricted least squares estimates because of the lack of a suitable distribution theory. Figure 1 gives a plot of the posterior means and the least squares estimates for the weights α_0 to α_8.

The first point to note from the results in Table 1 and Figure 1 is the erratic nature of the OLS weights. They are negative for lags 1 and 2, and then increase sharply to a maximum at lag 4, after which they decline, then increase, and then decline again. Forcing the weights to lie on a straight line produces estimates which are almost constant over the eight lags, declining only very slowly. The inequality restricted LS estimates (not plotted in Figure 1) take the form of a step function, being constant for 6 lags after the initial impact, and then declining. Using the Bayesian posterior means produces lag weights with a sharper, more credible decline. The initial impact is greater, and the final impact less, than the other estimates. After the first two lags, the decline is almost linear.

The marginal posterior pdfs for the lag weights are graphed in Figure 2. Note that, as the lag length increases, the location of the posterior pdfs for the weights at each lag length declines, and the precision of the estimation increases.

4. A LAG THAT INCREASES THEN DECLINES

For an example when the lag weights are expected to increase and then decline, we chose the Almon (1965) data where capital expenditures y_t are related to capital appropriations x_t in the current and preceding periods. Following the lead of a popular textbook (Hill, Griffiths and Judge, 1997), we choose a lag of 8, implying the model

$$y_t = \beta_0 + \alpha_0 x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \ldots + \alpha_8 x_{t-8} + e_t$$

Using inequality constraints to force the lag weights to increase and then decline means specifying one or two lags at which the weights are expected to reach a maximum. We estimated the following two cases:

1. The maximum lag weight is specified as α_4.
2. The maximum lag weight is α_3 or α_4, with the choice between these two being made by the data.

These choices were motivated by previously estimated lag schemes. The maximum is at α_4 if a polynomial distributed lag of degree 2 is estimated. It is at α_3 when estimated by unrestricted least squares.

Denoting the two feasible regions for the two cases that we considered by R_1 and R_2, we have

$$R_1 = \{\theta \mid 0 \le \alpha_0 \le \alpha_1 \le \alpha_2 \le \alpha_3 \le \alpha_4 \ge \alpha_5 \ge \alpha_6 \ge \alpha_7 \ge \alpha_8 \ge 0\}$$

$$R_2 = \{\theta \mid 0 \le \alpha_0 \le \alpha_1 \le \alpha_2 \le \alpha_3, \quad \alpha_4 \ge \alpha_5 \ge \alpha_6 \ge \alpha_7 \ge \alpha_8 \ge 0\}$$

For the purpose of inequality restricted least-squares estimation, the inequality restrictions for these cases can be written, respectively, as

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \\ \alpha_7 \\ \alpha_8 \end{bmatrix} \ge \begin{bmatrix} -\infty \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \\ \alpha_7 \\ \alpha_8 \end{bmatrix} \geq \begin{bmatrix} -\infty \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

The various estimated lag schemes are reported in Table 2. Like in the previous example, the Bayesian estimates are the posterior means; the values in parentheses are standard errors for sampling theory results and posterior standard deviations for the Bayesian results. The polynomial lag of degree 2 is described as a quadratic lag. Also included with the Bayesian estimates is the inequality restricted estimator that insists the lag weights are positive, but impose no other restrictions. These estimates are slightly different from the OLS estimates which, from a Bayesian standpoint, still allocate some positive probability to negative values.

A number of pairwise comparisons of different estimated lag schemes are graphed in Figures 3-6. Pairwise comparisons are used to emphasize particular differences and to avoid the congestion that occurs when too many lag schemes are drawn on the one figure. From Table 2 and Figures 3-6, we can make the following observations:

1. The OLS-estimated lag scheme conflicts with a priori expectations at lags 7 and 8. The slight increase at lag 7, and the larger increase at lag 8 do not conform with the view that the weights decline towards the end of the lag distribution.

2. The quadratic lag conforms with expectations. However, it produces a lag distribution that is much flatter than that from OLS, with the large weights at lags 2, 3 and 4 reduced considerably. (See Figure 3.)

3. When compared with OLS, the Bayesian inequality restricted estimator (with maximum at 3 or 4) also reduces large weights and increases small weights. (See Figure 4.) However, the effect is much less dramatic than with the quadratic lag estimator. The maximum weight still occurs at lag 3 and is similar in magnitude to that from OLS. Thus, the Bayesian inequality restricted estimator provides a lag scheme which conforms with our prior views, and conflicts less with the information provided by the data.

4. If we force the maximum weight to occur at lag 4 instead of letting the data choose between 3 and 4, the essential shape of the lag distribution is similar, except the maximum moves from 3 to 4. (See Figure 5.)

5. A comparison of the quadratic lag with the Bayesian inequality restricted with a maximum at lag 4 (Figure 6) reveals very similar lag distributions, except the greater flexibility of the Bayesian estimator has led to a larger impact from the largest weight.

6. The inequality restricted ML estimator that sets the maximum at lag 3 or 4 gives a reasonable flexible lag distribution until lag 5, after which the estimated lag weights are constant. When the maximum is set at lag 4, the weights at lags 2, 3 and 4 also become constant.

As examples of posterior densities for some of the lag weights, those for α_0, α_3 and α_5 are graphed in Figures 7 and 8 for the two different Bayesian estimated lag schemes. What is particularly noticeable from a comparison of the two figures is the difference in the posterior densities for α_3. Forcing the maximum to be at lag 4 implies greater precision in our knowledge about α_3. When the maximum can be at lag 3 or 4, α_3 can take on a much wider range of possible values.

5. CONCLUDING REMARKS

Bayesian estimation with inequality restrictions provides a method for including typical distributed-lag prior information without the need to force lag weights to lie on a less flexible continuous function. We have shown how the approach can be implemented for continuously declining lags, and for lags that increase and then decline. Our approach has two restrictions which could be relaxed in future research. The first is the need to specify, a priori, one or two lags at which a lag distribution reaches a maximum. A researcher may be reluctant to be so restrictive about where the maximum lies. The second restriction is a priori specification of the finite lag length. The relaxing of both the restrictions could be investigated within the context of a model averaging framework (Geweke, 1999) where alternative points for a maximum lag weight, and alternative lag lengths, could be introduced.

Table 1: Restricted Least Squares and Bayesian Parameter Estimates

Parameter	OLS	Equality Restricted LS	Inequality Restricted LS	Bayesian
β_1	-0.0534 (0.0084)	-0.0544 (0.0071)	-0.0529	-0.0509 (0.0054)
β_2	0.3602 (0.0463)	0.3740 (0.0444)	0.3804	0.3369 (0.0343)
α_0	0.0297 (0.0212)	0.0152 (0.0087)	0.0320	0.0377 (0.0099)
α_1	-0.0080 (0.0197)	0.0147 (0.0066)	0.0121	0.0252 (0.0052)
α_2	-0.0041 (0.0197)	0.0143 (0.0046)	0.0121	0.0204 (0.0041)
α_3	0.0177 (0.0189)	0.0138 (0.0030)	0.0121	0.0170 (0.0037)
α_4	0.0377 (0.0201)	0.0133 (0.0026)	0.0121	0.0140 (0.0036)
α_5	0.0098 (0.0191)	0.0128 (0.0038)	0.0121	0.0108 (0.0034)
α_6	0.0245 (0.0198)	0.0124 (0.0057)	0.0121	0.0080 (0.0032)
α_7	0.0118 (0.0204)	0.0119 (0.0078)	0.0119	0.0051 (0.0027)
α_8	0.0065 (0.0213)	0.0114 (0.0099)	0.0032	0.0025 (0.0020)
$\sum \alpha_i$	0.1256 (0.0248)	0.1199 (0.0233)	0.1197	0.1407 (0.0216)

Table 2: Lag Coefficients for Almon Data

	OLS	Quadratic Lag	Inequality Restricted *ML*: Max'm is α_4	Inequality Restricted *ML*: Max'm is α_3 or α_4	Bayesian Inequality Restricted: $\alpha_i > 0$	Bayesian Inequality Restricted: Max'm is α_4	Bayesian Inequality Restricted: Max'm is α_3 or α_4
α_0	0.038 (0.035)	0.067 (0.015)	0.046	0.044	0.043 (0.023)	0.046 (0.018)	0.038 (0.018)
α_1	0.067 (0.069)	0.100 (0.005)	0.074	0.066	0.075 (0.045)	0.090 (0.018)	0.089 (0.023)
α_2	0.181 (0.089)	0.123 (0.005)	0.168	0.177	0.173 (0.063)	0.129 (0.017)	0.139 (0.025)
α_3	0.194 (0.093)	0.136 (0.009)	0.168	0.198	0.175 (0.077)	0.154 (0.018)	0.197 (0.037)
α_4	0.170 (0.093)	0.138 (0.011)	0.168	0.129	0.168 (0.083)	0.188 (0.023)	0.148 (0.028)
α_5	0.052 (0.092)	0.130 (0.009)	0.078	0.080	0.061 (0.042)	0.120 (0.021)	0.109 (0.016)
α_6	0.052 (0.094)	0.112 (0.005)	0.078	0.080	0.068 (0.055)	0.090 (0.015)	0.091 (0.013)
α_7	0.056 (0.094)	0.083 (0.007)	0.078	0.080	0.087 (0.048)	0.068 (0.016)	0.071 (0.014)
α_8	0.127 (0.060)	0.044 (0.018)	0.078	0.080	0.098 (0.027)	0.036 (0.018)	0.044 (0.021)
$\sum \alpha_i$	0.939 (0.012)	0.933 (0.011)	0.936	0.934	0.948 (0.012)	0.923 (0.011)	0.926 (0.013)

Figure 1: Estimated Lag Weights for Orange Data

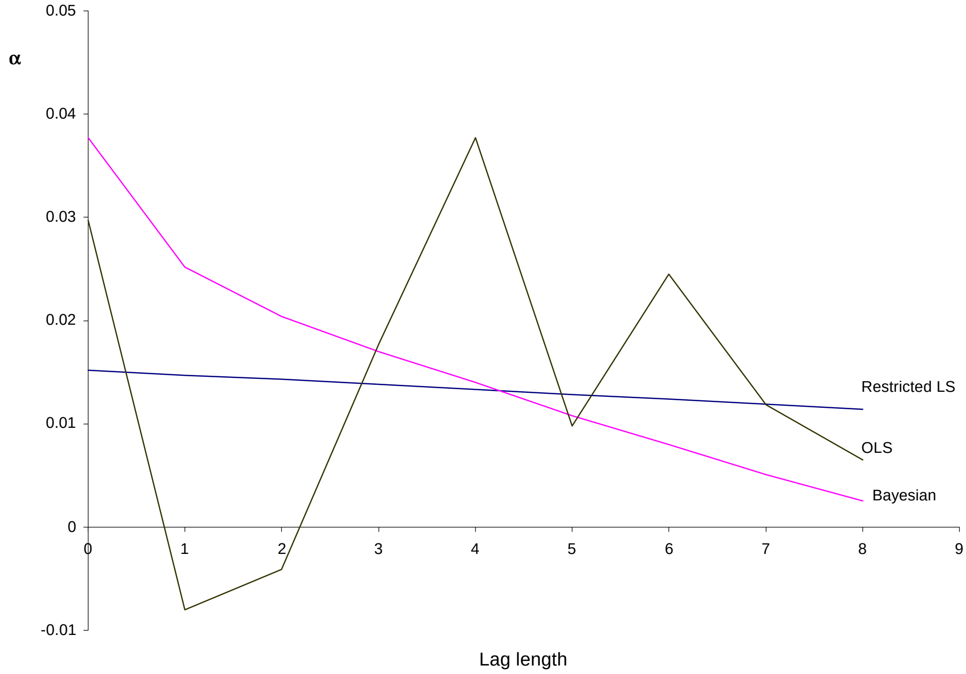


Figure 2: Marginal Posterior pdfs for the Lag Weights for Orange Data

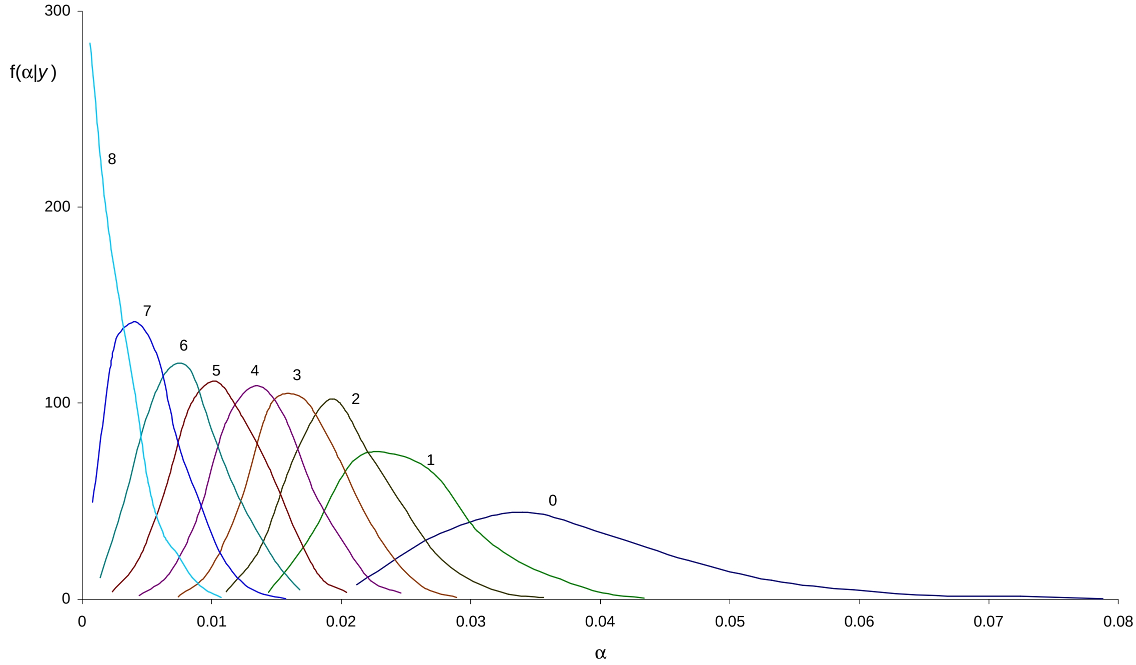


Figure 3: Lag Weights for Almon Data.
OLS and Quadratic Lag

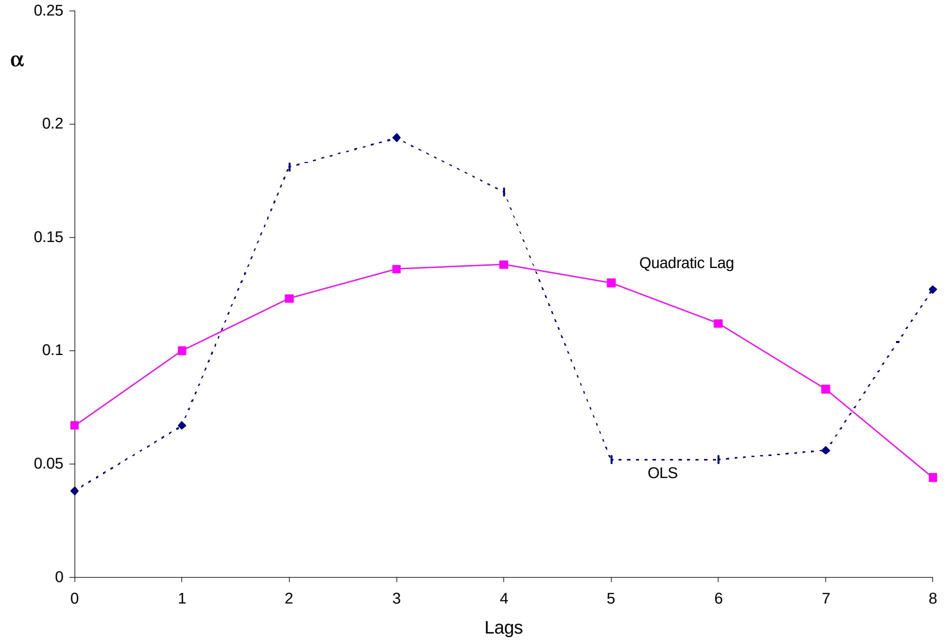


Figure 4: Lag Weights for Almon Data.
OLS and Bayes Inequality Restricted with Maximum at Lag 3 or 4

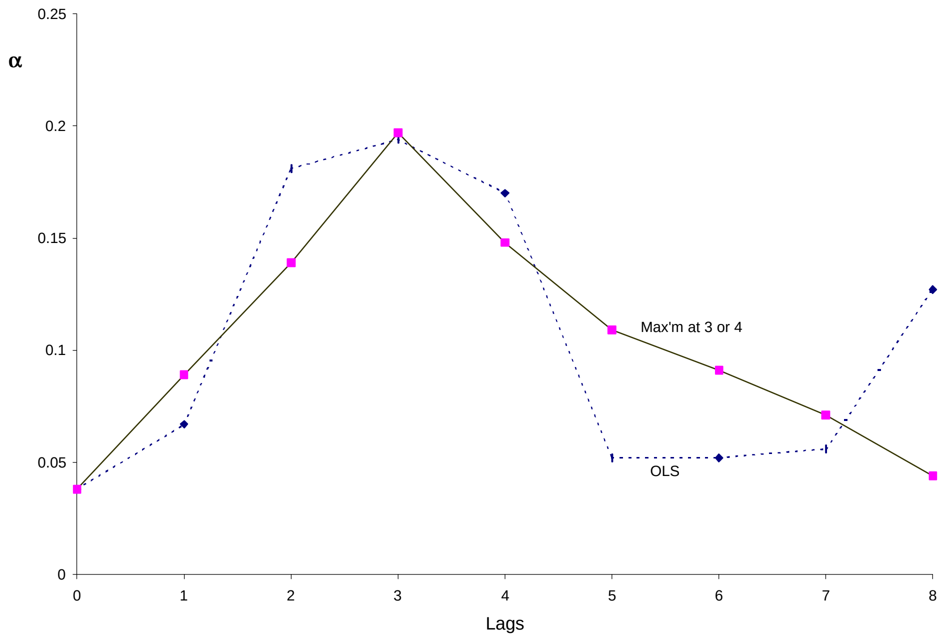


Figure 5: Lags Weights for Almon Data.
Bayes Inequality Restricted with Maximum at Lag 4 and
with Maximum at Lag 3 or 4.

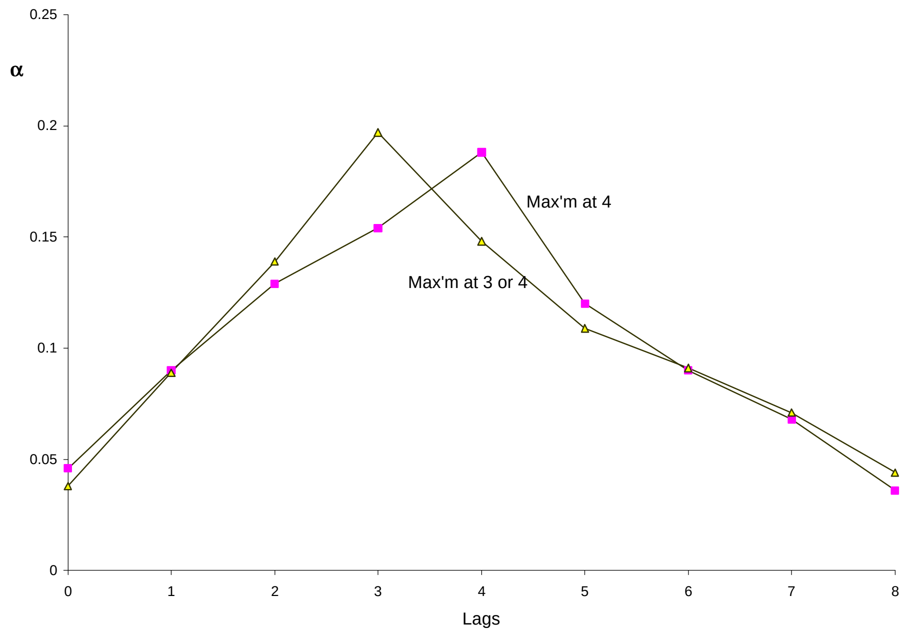


Figure 6: Lag Weights for Almon Data.
Quadratic Lag and Bayes Inequality Restricted with Maximum at Lag 4.

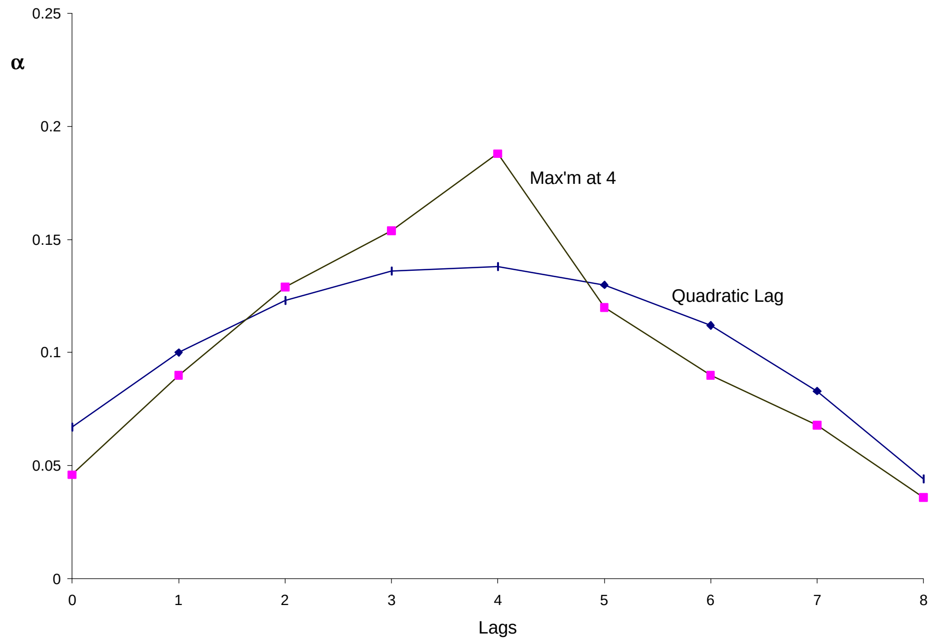


Figure 7: Posterior Distributions for Lag Weights with Maximum at Lag 4

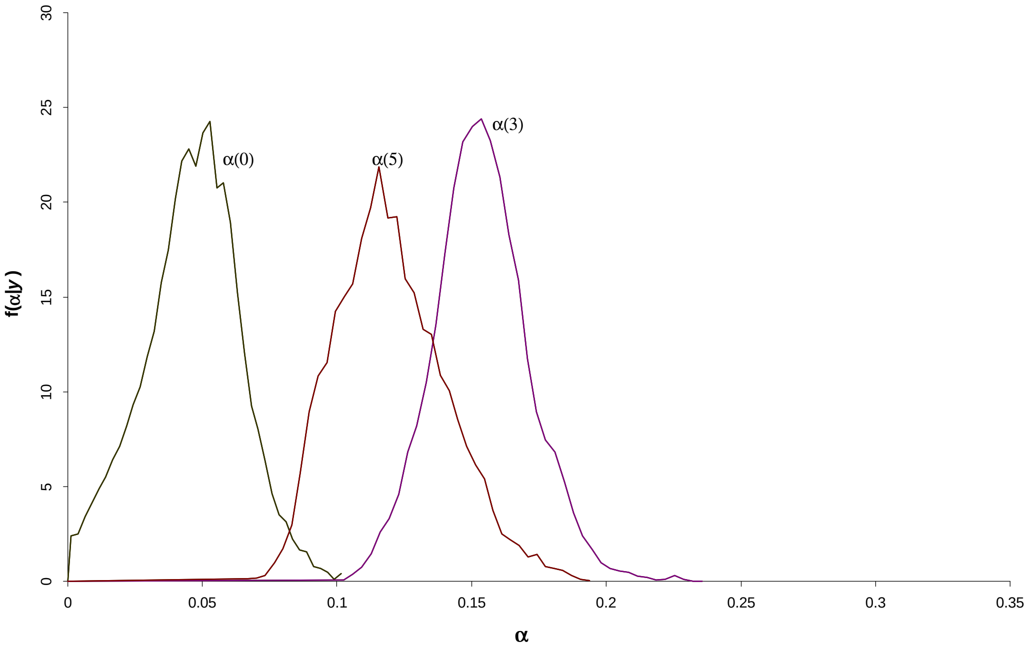
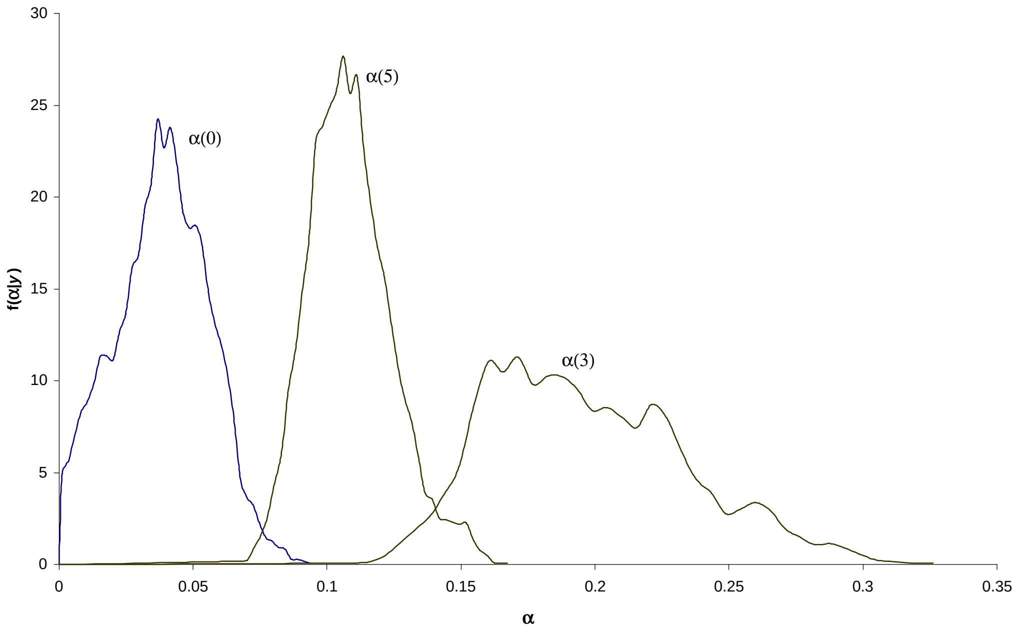


Figure 8: Posterior Distributions for Lag Weights with Maximum at Lag 3 or 4



REFERENCES

Albert, J.H. and Chib, S. (1996), "Computation in Bayesian Econometrics: An Introduction to Markov Chain Monte Carlo", in R.C. Hill (ed.), *Advances in Econometrics Volumn 11A: Bayesian Computational Methods and Applications*, JAI Press, Greenwich.

Almon, S. (1965), "The Distributed Lag between Capital Appropriations and Expenditures," *Econometrica*, 33, 178-196.

Baltagi, B.H. (1999), *Econometrics*, second edition, Springer, Berlin.

Berndt, E.R. (1991), *The Practice of Econometrics: Classic and Contemporary*, Reading, Addison Wesley.

Chib, S. and E. Greenberg (1996), "Markov Chine Monte Carlo Simulation Methods in Econometrics", *Econometric Theory*, 12, 409-431.

Davidson, R. and R. G. MacKinnon (1993), *Estimation and Inference in Econometrics*, Oxford University Press, New York.

Dhrymes, P. J. (1971), *Distributed Lags: Problems of Estimation and Formulation*, Holden-Day, San Francisco.

Dufour, J. and J.F. Kiviet (1998), "Exact Inference Methods for First-Order Autoregressive Distributed Lag Models," *Econometrics*, 66, 79-104.

Geweke, J. (1986), "Exact Inference in the Inequality Constrained Normal Linear Regression Model," *Journal of Applied Econometrics*, 1, 127-142.

Geweke, J. (1991), "Efficient Simulation from the Multivariate Normal and Student-*t* Distributions Subject to Linear Constraints, " in E.M. Keramidas (ed.), *Computing Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface*, Fairfax: Interface Foundation of North America, Inc. 571-578.

Geweke, J. (1996), "Bayesian Inference for Linear Models Subject to Linear Inequality Constraints," in W.O. Johnson , J.C. Lee and A. Zellner (eds.), *Modelling and Prediction: Honoring Seymour Geisser*, Springer-Verlag, New York, 248-263.

Geweke, J. (1999), "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication," *Econometric Reviews*, 18, 1-73.

Gilks, W.R., S. Richardson and D.J. Spirgelhalter (1996), *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London.

Greene, W.H. (1997), *Econometric Analysis*, third edition, Prentice-Hall, New Jersey.

Griffiths, W.E. and D. Chotikapanich (1997), " Bayesian Methodology for Imposing Inequality Constraints on a Linear Expenditure System with Demographic Factors," *Australian Economic Papers*, 36, 321-341.

Griliches, Z. (1967), “Distributed Lags: A Survey”, *Econometrica*, 35, 16-49.

Hendry, D.F., A.R. Pagan and J.D. Sargan (1984), “Dynamic Specification,” in Z.Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics*, Vol. 2, North-Holland, Amsterdam, 1025-1100.

Hill, R.C., W.E. Griffiths and G.G. Judge (1997), *Undergraduate Econometrics*, John Wiley & Sons, New York.

Jorgenson, D. W. (1965), “Anticipations and Investment Behavior”, in J. S. Duesenberry, G. Fromm, L. R. Klein and E. Kuh (eds.), *The Brookings Quarterly Econometric Model of the United States,* , Rand McNally & Company, Chicago, 35-94.

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

Judge, G.G., R. C. Hill, W.E. Griffiths, H. Lutkepohl and T. Lee (1985), *The Theory and Practice of Econometrics*, second edition, John Wiley & Sons, New York.

Judge, G.G. and T. Takayama (1966), “Inequality Restrictions in Regression Analysis,” *Journal of the American Statistical Association*, 61, 166-181.

Judge, G.G. and T.A. Yancey (1986), *Improved Methods of Inference in Econometrics*, North-Holland, Amsterdam.

Koyck, L. M. (1954), *Distributed Lags and Investment Analysis*, North-Holland, Amsterdam.

Lutkepohl, H. (1981), “A Model for Non-negative and Non-positive Distributed Lag Functions,” *Journal of Econometrics*, 16, 211-220.

Nerlove, M. (1956), “Estimates of the Elasticities of Supply of Selected Agricultural Commodities,” *Journal of Farm Economics,* 38, 496-509.

Nerlove, M. (1958), *The Dynamics of Supply: Estimation of Farmers' Response to Price*, The Johns Hopkins Press, Baltimore.

Nerlove, M. and F. Waugh (1961), "Advertising Without Supply Control: Some Implications of a Study of the Advertising of Oranges,” *Journal of Farm Economics*, 43, 813-837.

Tanner, M.A. (1993), *Tools for Statistical Inference*, Springer-Verlag, New York.

Toro-Vizcorrondo, C. and T. D. Wallace (1968), "A Test of the Mean Square Error Criterion for Restrictions in Linear Regression," *Journal of the American Statistical Association*, 63, 558-572.