Lag-Length Selection Criteria: Empirical Results from the St. Louis Equation

Authors	Dallas S. Batten, and Daniel L. Thornton
Working Paper Number	1983-008A
Citable Link	https://doi.org/10.20955/wp.1983.008
Suggested Citation	Batten, D.S., Thornton, D.L.; Lag-Length Selection Criteria: Empirical Results from the St. Louis Equation, Federal Reserve Bank of St. Louis Working Paper 1983-008. URL https://doi.org/10.20955/wp.1983.008

Published In	Federal Reserve Bank of St. Louis Review

LAG-LENGTH SELECTION CRITERIA: EMPIRICAL RESULTS FROM THE ST. LOUIS EQUATION

Dallas S. Batten & Daniel L. Thornton
Federal Reserve Bank of St. Louis
83-008

LAG-LENGTH SELECTION CRITERIA: EMPIRICAL RESULTS FROM THE ST. LOUIS EQUATION

Dallas S. Batten and Daniel L. Thornton*
Federal Reserve Bank of St. Louis
P. O. Box 442
St. Louis, MO 63166

This article describes and compares six criteria for determining the lag length of finite distributed lag models. These criteria are employed to select the lag length of the distributed lag variables within the St. Louis equation using a computationally efficient procedure. The lag lengths chosen are tested against each other and against arbitrarily overfitted and underfitted specifications. The results suggest that Akaike's final prediction error criterion and Pagano and Hartley's procedure perform well relative to the other criteria considered.

*Economists, Federal Reserve Bank of St. Louis. The views expressed are those of the authors and may not reflect those of the Federal Reserve Bank of St. Louis or the Board of Governors of the Federal Reserve System. The authors would like to thank Sarah Driver for her research assistance.

ID 6260R
DSB:DLT:cla
Disk 139R
Draft 9
6/30/83

Lag Length Selection Criteria: Empirical Results From the St. Louis Equation

1. INTRODUCTION

Recently, there has been an increased interest in lag-length selection techniques and criteria for determining the lag length of finite distributed lag models, e.g. Sargan (1980), Pagano and Hartley (1981), Geweke and Meese (1981) and Schwarz (1978). This interest is due, in part, to the increasingly widespread use of vector autoregression (VAR) models, such as Sims (1982) and Litterman (1982) and of the increased use of causality tests introduced by Granger (1969), Sims (1972) and Sargent (1976). VAR models and causality tests require that theoretically infinite distributed lags be represented by finite order regressions. Since the specification of the VAR model and the test of casuality are based on some classical hypothesis testing procedure, such as an F-test or a likelihood ratio test, they are dependent on choosing the "correct" lag length. Consequently, it is important that the lag length be specified carefully.

A number of different criteria have been suggested for determining the lag length. For example, recently Hsiao (1981) and Fey and Jain (1982) have suggested using Akaike's (1969,

1970) final prediction error (FPE) for determining the lag length for causality tests. Other criteria which have been suggested include Mallows' (1973) Cp-statistic, Schwarz' (1978) Bayesian Information Criterion (SBIC), the Bayesian Estimation Criterion (BEC) recently suggested by Geweke and Meese (1981) and standard tests for model specification, such as an F-test or a likelihood ratio test.

It is the purpose of this paper to compare the above criteria for selecting the "appropriate" lag length within the context of the St. Louis equation. The selected specifications are compared with each other and with arbitrarily chosen overfitted and underfitted specifications via likelihood ratio tests.

The organization of the paper is as follows. The various lag length selection criteria are presented and discussed in the second section. The empirical results of an application of each of these criteria are presented in the third section. Comparisons of the selected lags are made in the fourth section. A summary and conclusions are presented in the fifth section.

2. LAG-LENGTH SELECTION CRITERIA

Pagano-Hartley t-tests

Pagano and Hartley (1981) have developed a technique for selecting the appropriate lag length. To illustrate this procedure, consider the following general distributed lag model:

(1) $Y_t = \underline{\mu}\, \underline{Z}_t + \beta_\ell(L)X_t + \varepsilon_t \quad t=1, 2, \ldots, T,$

where $\varepsilon_t \sim NID(0, \sigma^2)$, $\underline{Z}'_t$ and $\underline{\mu}$ are k by 1 vectors of the nondistributed lag variables and parameters, respectively, $\beta_\ell(L)$ is the usual polynomial lag operator of order ℓ and X_t is an independent variable which affects Y_t with a lag of unknown length, ℓ.

The PH technique can be illustrated by rewriting equation (1) as

(2) $\underline{Y} = Z\underline{\mu} + X\underline{\beta} + \underline{\varepsilon}$,

where Z and X are T by k and T by (ℓ+1) matrices of observations on the independent variables, and $\underline{\mu}$ and $\underline{\beta}$ are k by 1 and (ℓ+1) by 1 vectors of parameters. The procedure begins by choosing a maximum lag length L. Equation (2), with the maximum lag length, can be rewritten as

(3) $\underline{Y}_L = W_L\, \underline{\psi}_L + \underline{\varepsilon}_L$,

where $W_L = [Z:X_L]$, and $\underline{\psi}_L = [\underline{\mu}:\ \underline{\beta}_L]'$.

The observation matrix W_L is then decomposed to

$$W_L = Q_L N_L$$

by the Gram-Schmidt decomposition. Here, Q_L is a matrix whose columns form an orthonormal basis for the column space of W_L and N_L is an upper triangular matrix with positive diagonal elements. Equation (3) now can be rewritten as

$$\underline{Y}_L = Q_L\, \underline{\lambda}_L + \underline{\varepsilon}_L,$$

where

$$\underline{\lambda}_L = [\underline{\lambda}^{\mu} : \underline{\lambda}_L^{\beta}]' = N_L \underline{\psi}_L .$$

Given that Q_L is orthonormal, the least squares estimate of $\underline{\lambda}_L$ is given by

$$\hat{\underline{\lambda}}_L = [\hat{\underline{\lambda}}^{\mu} : \hat{\underline{\lambda}}^{\beta}]' = Q_L ' \underline{Y}_L ,$$

and the structural parameters can be obtained from

$$N_L \hat{\underline{\psi}}_L = \hat{\underline{\lambda}}_L .$$ [1/]

Pagano and Hartley suggest that ℓ be specified by choosing the smallest j for which the null hypothesis, H_{L-j}: $\lambda^{\beta}_{L-j}=0$, is rejected.[2/]

While the Pagano-Hartley (PH) technique is an alternative criterion to those suggested above, it has an additional benefit of enabling the calculation of the various lag-length selection statistics in a computationally efficient manner.[3/]

The Standard F-Test

One procedure is to calculate a sequential F-statistic,

$$F_{L-j} = (RSS_{L-j-1} - \underline{e}_L' \, \underline{e}_L)/(j+1)s^2, \quad j=0, 1, 2, \ldots, L,$$

and select the lag length as the first L-j for which the null hypothesis, $\beta_L = \beta_{L-1} = \ldots = \beta_{L-j} = 0$, is rejected. PH note that the RSS_{L-j-1} (the sum of squared errors when j+1 restrictions are imposed) can be conveniently expressed in terms of the coefficients of the orthogonal regression, so that the above F-statistic can be written as

$$F_{L-j} = \sum_{k=L-j}^{L} (\hat{\lambda}^{\beta}_k)^2/(j+1)s^2 .$$ [4/]

Mallows' Cp-Statistic

An alternative to the above classical approaches is to consider minimizing some function of the residual sums of squares. One such statistic is Mallows' Cp-statistic, which is based upon a mean square error prediction norm. The Cp-statistic is defined as

$$Cp_{L-j} = \frac{1}{s^2} RSS_{L-j} - T + 2(L+K+1-j), \quad j=0,\ 1,\ 2,\ \ldots\ L.$$

As j increases from zero to L, the Cp-statistic trades off some reduction in "variance" of prediction for an increase in the "bias". The value of j for which the Cp-statistic is a minimum is the one that minimizes the expected mean square error of prediction. One problem with the Cp-statistic is that it will attain a local minimum whenever the t-statistic on the marginal distributed lag variable is greater than or equal to $\sqrt{2}$.[5/] This suggests a low power of discrimination.

Akaike's FPE Criterion

Another criterion based on a mean square error prediction norm is Akaike's FPE criterion defined as:

$$FPE_{L-j} = \frac{T + (L+K+1-j)}{T - (L+K+1-j)} \cdot \frac{RSS_{L-j}}{T}, \qquad j = 0,\ 1,\ .\ .\ .,\ L.$$

Like Mallows' Cp-statistic, the FPE criterion attempts to balance the "risk" due to bias when shorter lag lengths are selected against the "risk" due to the increase in variance when longer lag lengths are chosen. Choosing lag lengths by minimizing the FPE is equivalent to applying an approximate sequential F-test with varying significance levels.[6/]

Bayesian Criteria

Two Bayesian criteria have been suggested that select the correct lag length asymptotically. The first of these is Schwarz' Bayesian Information Criterion (SBIC) and is given by

$$SBIC = \ln \left[\frac{RSS_{L-j}}{T-L-K-1+j}\right] + (L+K+1-j)\frac{\ln (T)}{T}.$$

The second, suggested by Geweke and Meese, is the Bayesian Estimation Criterion (BEC), given by

$$BEC = \frac{RSS_{L-j}}{T-L-K-1+j} + (L+K+1-j)s^2 \left(\frac{\ln (T)}{T-L-K-1}\right).$$

These Bayesian criteria select the correct lag length asymptotically; their only advantage, however, is asymptotic efficiency, since choosing ℓ which is too long does not result in biased estimates of the distributed lag parameters.[7/]

3. APPLICATION TO THE ST. LOUIS EQUATION

To investigate the appropriate lag lengths for the St. Louis equation, we employ the now standard growth rate specification,

$$(4) \qquad \dot{Y}_t = \alpha + \sum_{i=0}^{J} \beta_i \dot{M}_{t-i} + \sum_{i=0}^{K} \gamma_i \dot{G}_{t-i} + \epsilon_t,$$

where the dots over each variable represent quarter-to-quarter annualized rates of change and Y, M and G are nominal GNP, money (the M1 definition) and high-employment government expenditures, respectively. The sample period considered is II/1962 to III/1982.

Each of the criteria for determining the appropriate lag lengths is employed to select the lag lengths (J, K) for money and government expenditures growth.[8/] To assess the sensitivity of the various techniques to the selection of the maximum lag length (L), three values of L (8, 12 and 16) are specified initially for each variable. The results are summarized in table 1.[9/]

The results from table 1 show that the chosen lag lengths differ by criterion and by maximum lag length specified. In general, the longest lags are selected by the PH t-test and Akaike's FPE criterion, while the shortest lags are obtained by the Bayesian criteria. The Bayesian criteria yield lags which appear to be too short *a priori*. Since they are designed to choose the appropriate lag length in the limit, it appears they may give too much weight to this quality in finite samples. These criteria, however, appear to be insensitive to the choices of L.

Mallows' Cp-statistic is also insensitive to the choice of L.[10/] However, for $L > 8$, it selects a lag length shorter than either the PH or the FPE criterion.[11/]

Not surprisingly, the F-test is the most sensitive to the choice of the maximum lag length. The lag length selected is the least restricted model for which the null hypothesis was rejected at the 5 percent level. The F-test tends to indicate shorter lags whenever the first significant lag coefficients are proceeded by a number of insignificant ones. These insignificant coefficients tend to dilute the discriminating power of the F-test. This is why the F-test chooses a much shorter lag when L is increased from 12 to 16. Of course, this is to be expected, given the general nature of the test.

The PH and FPE criteria select identical lag lengths for L = 12 or L = 16, and nearly identical lag lengths for L = 8.[12/] Furthermore, they appear to be insensitive to the choice of L as long as the maximum lag length considered is sufficiently large. The FPE norm, however, is sensitive to the order in which one proceeds. If the FPE criterion is employed with two or more distributed lag variables, care must be taken to insure that a global minimum is achieved.

4. TESTS OF ALTERNATIVE LAG STRUCTURES

As a diagnostic check, each lag structure in table 1 is tested against the others and against arbitrarily chosen lags of 4, 6 and 12 on both $\dot{M}$ and $\dot{G}$ using a likelihood ratio test. Because some of the lag structures reported in table 1 differ only slightly from each other, the results of all the tests are not reported in table 2.[13/] The results indicate that of all the lag structures, only the longer lags chosen by the PH and FPE criteria cannot be rejected relative to all of the other specifications. Furthermore, the extremely short lags of the Bayesian criteria are always rejected relative to the other selected lag specifications. While no amount of testing can be conclusive, these results suggest that the longer lag structures selected by the PH and FPE criteria are the most appropriate.

5. CONCLUSIONS

Our analysis of alternative lag-length selection criteria applied to the St. Louis equation suggests that both the Pagano and Hartley procedure and Akaike's final prediction error perform well in selecting the appropriate lag structure relative to the other techniques considered. While the scope of the experiment is limited, our results broadly support the recent findings of Hsiao (1981) with respect to the FPE criterion. Care must be exercised, however, to insure a global minimum of the FPE if two or more distributed lag variables are present.

Also, the orthogonal regression procedure, used for the Pagano and Hartley criteria, provides a computationally efficient method of calculating the statistics of the other criteria, including the FPE. This enables a convenient consideration of several criteria for determining the lag structure.

Finally, the procedure and criteria considered here can be extended easily to the polynomial degree selection of a polynomial distributed lag model.[14]

FOOTNOTES

1/One of the advantages of the PH method is that the elements of $\hat{\underline{\lambda}}_L$ are mutually independent random variables. In particular,

$$\hat{\lambda}_i^{\beta} \sim \text{NID}\ (\lambda_i,\ \sigma^2),\ i=0,\ 1,\ 2,\ \ldots,\ \ell \quad \text{and}$$

$$\hat{\lambda}_i^{\beta} \sim \text{NID}\ (0,\ \sigma^2),\ i=\ell+1,\ \ell+2,\ \ldots,\ L.$$

2/This hypothesis can be tested by employing a simple t-test of the following form:

$$t_{L-j} = \hat{\underline{\lambda}}^{\beta}_{L-j}/s, \qquad j=0,\ 1,\ 2,\ \ldots,\ L,$$

where

$$s^2 = \frac{\underline{e}_L'\ \underline{e}_L}{T-K-L-1}\ .$$

Pagano and Hartley conclude erroneously that the hypothesis,

$$H^*_{L-j}:\ \beta_L = \beta_{L-1} = \ldots = \beta_{L-j} = 0,\ j = 0,\ 1,\ \ldots,\ L,$$

is equivalent to H_{L-j}. For a discussion of the differences between these two hypotheses, see Batten and Thornton (1983a, pp. 21-22).

3/Specifically,

$$RSS_{L-j-1} = Y_L'Y_L - \sum_{k=1}^{K} (\hat{\lambda}_k^{\mu})^2 - \sum_{k=0}^{L-j-1} (\hat{\lambda}_k^{\beta})^2,\ j = 0,\ 1,\ 2,\ \ldots,\ L.$$

4/F_{L-j} is distributed as an F with (j+1) and (T-L-K-1) degrees of freedom.

5/See Sawyer (1979). It can be shown that if the regressors are orthogonal, the equality holds. That is, $\Delta Cp = (\text{PH t-statistic})^2 - 2$.

6/See Hsiao (1981).

7/For a discussion of the properties of these Bayesian criteria, see Geweke and Meese (1981) and Schwarz (1978).

8/These procedures are somewhat more complicated when appropriate lag lengths are selected for two or more variables. See Batten and Thornton (1983a) for details.

9/The full set of results is presented in Batten and Thornton (1983b).

10/Although minimizing the Cp-statistic is a commonly cited criterion, Judge, et. al., admonish its use in applied work. Instead, they suggest selecting Cp=K, where K is the number of included variables in the model. Unfortunately, in our case there were a number of lag specifications which satisfied this criterion approximately. See Judge, et. al., (1981, pp. 418-19).

11/The fact that the Cp-statistic selects a shorter lag then the FPE criterion is not too surprising. If we consider only two alternative lag specifications p and p+q the FPE criterion will pick p if $F < 2T/T+p+q+1$ and p+q if $F > 2T/T+p+q+1$ [see Hsiao (1981)]. Alternatively, the Cp-statistic will select p if $F < 2$ and p+q if $F > 2$. Thus, within the context of this limited framework, there will be a tendency for the Cp-statistic to select the shorter lag.

12/Actually, the PH t-ratio for the second lag of $\dot{G}$ was 1.91 when the lag on $\dot{M}$ was 5. Thus, the PH technique nearly selected 5 and 2 also.

[13]The complete set of test results is reported in Batten and Thornton (1983b).

[14]For example, see Batten and Thornton (1983a).

Table 1

Results for Various Lag-Length Selection Criteria

	Maximum Lag Length					
Criterion	8		12		16	
	$\dot{M}$	$\dot{G}$	$\dot{M}$	$\dot{G}$	$\dot{M}$	$\dot{G}$
PH t-test	5	0	10	9	10	9
Mallows' Cp	5	2	5	2	5	2
Akaike's FPE	5	2	10	9	10	9
BEC	1	0	1	0	1	0
SBIC	2	0	2	0	2	0
F-test	1	8	10	8	1	0

Table 2

Likelihood Ratio Test Results

Lags on		Lags on $\dot{M}$ and $\dot{G}$					
$\dot{M}$	$\dot{G}$	10 9	6 6	5 2	4 4	2 0	0 0
12	12	3.31	N.A.	29.90*	N.A.	43.53*	N.A.
10	9	--	21.26*	26.59*	32.36*	40.22*	--
6	6	--	--	5.34	N.A.	18.96*	N.A.
5	2	--	--	--	--	13.63*	--
4	4	--	--	--	--	7.86	N.A.
2	0	--	--	--	--	--	18.10*

N.A. indicates that the likelihood ratio was not calculated between arbitrarily chosen lags.

* Statistically significant at the 5 percent level.

References

Akaike, H. (1969), "Statistical Predicator Identification," Annals of the Institute of Statistical Mathematics, 21, 203-17.

___________ (1970), "Autoregressive Model Fitting For Control," Annals of the Institute of Statistical Mathematic, 22, 163-80.

Batten, D., and Thornton, D. (1983a), "Polynomial Distributed Lags and the Estimation of the St. Louis Equation," Federal Reserve Bank of St. Louis Review, 65, 13-25.

__________ and ___________ (1983b), "Complete Results for Lag Length Selection," Federal Reserve Bank of St. Louis Research Paper 83-009.

Fey, R. A., and Jain, N. C. (1982), "Identification and Testing of Optimal Lag Structures and Causality in Economic Forecasting," in Applied Time Series Analysis, O. D. Anderson and M. R. Perryman, eds., North-Holland, 65-73.

Geweke, J., and Meese, R. (1981), "Estimating Regression Models of Finite But Unknown Order," International Economic Review, 22, 55-70.

Granger, C. W. J. (1969), "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," Econometrica, 37, 424-38.

Hsiao, C. (1981), "Autoregressive Modelling and Money-Income Causality Detection," Journal of Monetary Economics, 7, 85-106.

Judge, G., and Griffiths, W., Hill, R., and Lee, T. (1980), The Theory and Practice of Econometrics, John Wiley and Sons, Inc.

Litterman, R. B. (1982), "Optimal Control of the Money Supply," Federal Reserve Bank of Minneapolis Working Paper 200.

Mallows, C. L. (1973), "Some Comments on Cp," Technometrics, 15, 661-75.

Pagano, M., and Hartley, M. (1981), "On Fitting Distributed Lag Models Subject to Polynomial Restrictions," Journal of Econometrics, 16, 171-98.

Sargent, T. J. (1976), "A Classical Macroeconometric Model for the United States," Journal of Political Economy, 84, 207-37.

Sargan, J. D. (1980), "The Consumer Price Equation in the Post War British Economy: An Exercise in Equation Specification Testing," The Review of Economic Studies, 47, 113-35.

Sawyer, K. R. (1979), "An Examination of Some Model Selection Criteria," mimeo, Australian National University.

Schwarz, G. (1978), "Estimating the Dimension of a Model," Annals of Statistics, 6, 461-64.

Sims, C. A. (1972), "Money, Income, and Causality," The American Economic Review, 62, 540-552

__________ (1982), "Policy Analysis with Econometric Models," Brookings Papers on Economic Activity, 1982, 107-52.