Weak Identification, the Pile-Up Problem, and Finite-Sample Inference for ARMA Models

James Morley
University of New South Wales Sydney NSW 2052 Australia

Charles R. Nelson
University of Washington Seattle WA 98195 USA

Irina B. Panovska
Lehigh University Bethlehem PA 18015 USA

Draft November 30, 2016

EXTENDED ABSTRACT

An ARMA(1,1) model is weakly identified when the AR and MA parameters are close to each other because it is their difference than determines the amount of information in the data. While point estimates are imprecise in that case, it is well established that estimated standard errors provided by standard asymptotic distribution theory are much too small, leading to an artificial sense of precision and massive size distortions of tests in finite samples.

Andrews and Cheng (2012) (hereafter A-C) have derived asymptotic distributions for parameters of weakly identified models and applied their results to the ARMA(1,1) model, but the assumptions of their theory require that the parameter space be restricted to well within the limits of stationarity for AR and invertibility for MA to 'avoid unit root and boundary effects'. However, values close to these limits are often reported in empirical research and may be of considerable economic interest; for example, many theoretical macroeconomic models with highly persistent shocks imply observable variables such as output and consumption growth follow ARMA and VARMA processes with AR and MA parameters close to +1. The prevalence of near boundary estimates for the MA parameter in particular is called the 'pile-up' problem and has been noted in numerous studies. Thus the need for an approach that encompasses the entire stationary/invertible parameter space is clear, and the objective of this paper is to provide one.

The origin of problems that occur near the stationary/invertibility boundary become clear when we examine the surface of the log likelihood (hereafter LL) for the ARMA model when the data are white noise. The AR and MA parameter are not identified in this case, and we would expect any single value for both in the range -1 to +1 to be as consistent with the data as is zero since the model then reduces by cancellation back to white noise. Thus, the surface will be dominated by a ridge of constant log likelihood. What one finds is that there is a region of the parameter space where the LL is even higher than along this ridge. For sample size n=100 it is around

the values .85 for the AR coefficient and 1.0 for the MA coefficient, and the gain in LL is about 2. The explanation for this is that when the MA coefficient is +1 the ARMA model is the time trend regression studied by Dickey and Fuller (1979) (hereafter D-F) after first differencing – levels of the random walk become white noise and the time trend becomes the intercept while the error becomes MA(1) with a unit root. We know from their theory and results that the AR coefficient in this regression will be estimated to be around .85 for n=100, which is what we observe here. In seven of ten samples the maximum of the LL occurs with the MA coefficient at +1 and at .9 in another, and within the restricted parameter space of A-C, [-.8, +.8], in only two. Estimates of the AR coefficient cluster around .85 and are within the A-C only for the same two samples. Thus, the D-F regression dominates the results of estimating the ARMA model when the data are simply white noise.

When we slice the LL surface along the fixed MA value +1 we have the LL for the lag coefficient in the D-F regression coefficient with its peak, as expected, around .85. The gain in log likelihood relative to the AR at +1 is about 2, which we know to not be evidence against the unit root hypothesis that it is +1 but rather that it is consistent with the unit root.

When we slice the LL surface at a fixed value for the AR coefficient of, say .8, we are seeing the log likelihood for the MA(1) model with true coefficient .8; that is readily seen by examining the sum of squares. What is revealed is that the LL reaches a peak not at a value of +.8 for the MA parameter but at +1. Thus, the well-known and widely documented pile-up phenomenon encountered in estimation of the MA(1) model is also explained by the D-F regression theory.

Clearly, the situation for practitioners is greatly improved if we can build on the A-C results to develop a strategy for inference and testing in the ARMA model which is robust to weak identification in the sense of providing correct, or more correct, confidence intervals and test size. In this paper we propose a modification of the A-C tests based on mixtures of distributions. Our proposed modified tests address both weak identification and the pile-up problem, outperforming the original Andrews-Cheng tests in finite samples.

The mixture distributions used for inference combine the A-C distribution with the D-F distribution for the AR parameter and the A-C distribution with a point mass at +1 for the MA parameter. We set the weights on the underlying distributions based on the probability of pile-up for the MA parameter. The pile-up probability depends on the true value of the MA parameter, which is consistently estimated when making inference about the AR parameter or is directly hypothesized when making inference about the MA parameter. Building on the results in Davis and Dunsmuir (1996) and Chen, Davis, and Song (2011), we estimate a lower-bound on the pile-up probability and we find in Monte Carlo simulations for a variety of data generating processes that tests for AR and MA parameters using our mixture distributions based on the estimated pile-up probability are robust for parameters well within the stationarity/invertibility bounds or under strong identification, but have better empirical size than when using the A-C distribution or a standard asymptotic

distribution in sample sizes of n=100 or n=250 for parameters close to the bounds and/or under weak identification.

JEL Classification: C1

MEG Keywords: Time Series, Macroeconometrics, Identification, Finite Sample Inference, Weakly Identified Models.