E C O N O M I C S B U L L E T I N

Markov chain approximation in bootstrapping autoregressions

Stanislav Anatolyev
New Economic School, Moscow

Andrey Vasnev
New Economic School, Moscow

Abstract

We propose a bootstrap algorithm for autoregressions based on the approximation of the data generating process by a finite state discrete Markov chain. We discover a close connection of the proposed algorithm with existing bootstrap resampling schemes, run a small Monte−Carlo experiment, and give an illustrative example.

Our thanks go to the associate editor Serena Ng, an anonymous referee, Sergey Ayvazyan, Oleg Zamulin, and the participants of the VIIth Spring Meeting of Young Economists at University of Paris 1 Panthéon−Sorbonne, 2002. Research for this note was supported in part by a fellowship from the Economics Education and Research Consortium, with funds provided by the Government of Sweden through the Eurasia Foundation. The opinions expressed in this note are those of the authors and do not necessarily reflect the views of the Government of Sweden, the Eurasia Foundation, or any other member of the Economics Education and Research Consortium.
Citation: Anatolyev, Stanislav and Andrey Vasnev, (2002) "Markov chain approximation in bootstrapping autoregressions." *Economics Bulletin,* Vol. 3, No. 19 pp. 1−8
Submitted: July 24, 2002. **Accepted:** September 17, 2002.
URL: http://www.economicsbulletin.com/2002/volume3/EB−02C10004A.pdf

1. Introduction

A good resampling procedure that generates bootstrap samples with dependence structure adequately representing the underlying data generating process (DGP), is a critical requirement for successful bootstrap inference with stationary time series data. Earlier proposals, like the residual bootstrap and block bootstrap, leave one dissatisfied for one reason or another. Recently there has been a movement towards nonparametric bootstrap algorithms that incorporate estimation of the conditional probability measure. Among these is the *Markov conditional bootstrap* (Rajarshi, 1990, Horowitz, 2001), where the conditional density is estimated by kernel methods and bootstrap data are generated by successively sampling observations according to the estimated density. In a recent paper, Paparoditis and Politis (2001a) proposed a resampling scheme called the *local bootstrap* (LB) that does not require estimation of multivariate densities and sampling from a continuous distribution (see also Paparoditis and Politis, 2001b). The authors motivate the LB by its being a direct extension of Efron's (1979) bootstrap for IID data to the case of stationary time series models with Markovian structure.

In this note, we propose a similar resampling algorithm, the *Markov chain bootstrap* (MCB), based on the approximation of the DGP by a finite state discrete Markov chain. Previous literature where Markov chain approximation was used includes Tauchen (1986) where it helped to numerically solve an integral equation, and Gregory (1989), where it helped to test for ARCH effects. It turns out that extending the MCB in a natural way yields the LB algorithm. Thus the MCB provides a convenient link for motivation of the LB. Also, we discover the relation of the MCB and LB to Hansen's (1999) nonparametric bootstrap designed for models with conditional moment restrictions. A simulation experiment shows that the MCB has attractive size properties in small samples. In an application to US GNP, the MCB inference leads to an acceptance of the null of no predictability within a simple class of models, while the asymptotic inference – to its strong rejection.

Paparoditis and Politis (2001a, section 3) show asymptotic validity of the local bootstrap for linear statistics of geometrically strong mixing Markov processes under smoothness and boundedness (away from zero on a compact set) conditions placed on the transition density. We conjecture that in similar circumstances the MCB is asymptotically valid too. There do not yet exist results on whether the LB achieves an asymptotic refinement, but the closely related Markov conditional bootstrap does deliver higher order improvements under suitable conditions relative not only to asymptotics, but also to the block bootstrap, as shown in Horowitz (2001, sections 3–4) for Markov and even approximately Markov processes that are geometrically strong mixing.

2. Markov chain bootstrap

The MCB resampling algorithm is based on the approximation of the unknown DGP by a finite state discrete Markov chain. Suppose that we have a (possibly nonlinear) stationary autoregression for variable y_t, with L lags present on the right hand side as vector y_t^-. Let us partition the support Ψ of y_t so that $\Psi = \bigcup_{k=1}^{I} \Psi_k$. This also induces partition of the support of the regressors y_t^-: $\Psi^- = \times_{\ell=1}^{L} \bigcup_{k_\ell=1}^{I} \Psi_{k_\ell}$. Let us also associate y with its bin $\Psi(y)$ in the partition, i.e. $\Psi(y) = \Psi_k$ for such k that $y \in \Psi(y)$. Similarly, let $\Psi^-(y^-) = \times_{\ell=1}^{L} \Psi_{k_\ell}$, where k_ℓ, $\ell = 1, \cdots, L$, are such that $y^- \in \Psi^-(y^-)$. Thus, any pair (y, y^-) is associated with its cell $\Psi(y) \times \Psi^-(y^-)$ in the partition.

We approximate the conditional density $f\left(y|y^-\right)$ by its discrete state space analog, an I^L by I transition matrix of the finite state discrete Markov chain filled with probabilities $\Pr\left(\Psi(y)|\,\Psi^-(y^-)\right)$. These probabilities are unknown but may be nonparametrically estimated from the sample whenever the cell $\Psi(y) \times \Psi^-(y^-)$ is not empty:

$$
\begin{aligned}
\widehat{\Pr}\left(\Psi(y)|\,\Psi^-(y^-)\right) &= \frac{\widehat{\Pr}\left(\Psi(y) \times \Psi^-(y^-)\right)}{\widehat{\Pr}\left(\Psi^-(y^-)\right)} \\
&= \frac{(T-L)^{-1} \sum_{t=L+1}^{T} \mathbb{I}\left(y_t \in \Psi(y)\right) \mathbb{I}\left(y_t^- \in \Psi^-(y^-)\right)}{(T-L)^{-1} \sum_{t=L+1}^{T} \mathbb{I}\left(y_t^- \in \Psi^-(y^-)\right)},
\end{aligned} \tag{1}
$$

where $\mathbb{I}(\cdot)$ is the indicator function.

Now we are ready to construct a bootstrap sample. The initial L bootstrap observations are jointly drawn randomly from the sample, or simply set equal to those in the sample. Consider $L < s \leq T$. Suppose that the part $y_1^*, \cdots, y_{s-1}^*$ of the bootstrap sample has been constructed. The next bootstrap observation y_s^* is drawn from the $(T-L)$-tuple $(y_{L+1}, \cdots, y_T)$ with probabilities $(\hat{q}_{L+1}, \cdots, \hat{q}_T)$, which we call the *resampling rule*, where each $\hat{q}_\tau$, $\tau = L+1, \cdots, T$, is an estimate of the following probability of selecting the candidate y_τ conditional on the bootstrap regressor:

$$
q_\tau = \mathbb{I}\left(\Psi^-(y_\tau^-) = \Psi^-(y_s^{-*})\right) \times \Pr\left(\Psi(y_\tau)|\,\Psi^-(y_\tau^-)\right) \times \left|\Psi(y_\tau) \times \Psi^-(y_\tau^-)\right|^{-1}, \tag{2}
$$

where $|\cdot|$ counts observations within a cell. The first term converts the regressors to the terms of bins in the partition. The second term is the transition probability of the designed

Markov chain corresponding to the candidate's cell. The third term converts the bins back to the terms of y. If the transition probability is estimated by (1), then

$$\hat{q}_\tau \propto \mathbb{I}\left(y_\tau^- \in \Psi^-(y_s^{-*})\right). \tag{3}$$

The number of observations within the cell drops out, and as a result we have a very simple resampling rule: the unit probability mass is distributed uniformly among those y_τ's in the sample whose regressors y_τ^- belong to the same bins as the bootstrap regressors y_s^{-*} do.

The indicator function in (3) may be viewed as a proximity measure based on a histogram kernel. If this measure employed a smooth resampling kernel, the MCB resampling rule would be that of the local bootstrap of Paparoditis and Politis (2001a, 2001b):

$$\hat{q}_\tau \propto K_h\left(y_\tau^- - y_s^{-*}\right), \tag{4}$$

where $K_h(\cdot)$ is a bandwidth-rescaled resampling kernel.

Also note that the resampling rule (4) solves the following optimization problem:

$$(\hat{q}_{L+1}, \cdots, \hat{q}_T) = \underset{(q_{L+1},\cdots,q_T)}{\arg\max} \left\{ \sum_{\tau=L+1}^{T} K_h\left(y_\tau^- - y_s^{-*}\right) \log(q_\tau) \quad \text{s.t. } q_\tau \geq 0, \sum_{\tau=L+1}^{T} q_\tau = 1 \right\}.$$

This is a problem of the local empirical likelihood estimation of the conditional distribution. Recently Hansen (1999) suggested a nonparametric bootstrap procedure based on empirical likelihood that is designed for inference in problems with a conditional moment restriction. Hansen's optimization problem for the resampling rule is as one above, except that it contains an additional constraint due to the moment restriction. This constraint distorts the resampling rule away from (4). Therefore, the MCB and LB may be viewed as special cases of Hansen's (1999) algorithm when applied in the absence of a conditional moment restriction.

3. Simulation evidence

In this small simulation experiment, we study the MCB inference about the slope of the linear projection of a stationary variable y_t on its own one-step-back past:

$$y_t = \mu + \alpha y_{t-1} + e_t, \tag{5}$$

where e_t is the projection error with properties $E[e_t] = E[y_{t-1}e_t] = 0$ that render the OLS estimators $\hat{\mu}$ and $\hat{\alpha}$ consistent. The autocorrelation properties of the error e_t depend on the true DGP generating y_t. We consider three alternative mechanisms that imply the true value of 0 for both μ and α:

$$\begin{aligned} \text{DGP A} &: \quad y_t = \varepsilon_t, \\ \text{DGP B} &: \quad y_t = \tfrac{1}{2} y_{t-2} + \varepsilon_t, \\ \text{DGP C} &: \quad y_t = \tfrac{1}{2} y_{t-2}\varepsilon_{t-1} + \varepsilon_t. \end{aligned}$$

In all of them, the true innovation ε_t is an ARCH-type heteroskedastic martingale difference sequence relative to the past and is generated as

$$\varepsilon_t = \eta_t \sqrt{1 + \tfrac{1}{2}\varepsilon_{t-1}^2}, \quad \eta_t \sim IID\ \mathcal{N}(0,1).$$

DGP A is an autoregression of order one, so e_t is the true innovation ε_t and thus is serially uncorrelated. DGP B is an autoregression of order 2, so e_t is an infinite linear distributed lag on the true error ε_t, and thus is serially correlated of infinite order. DGP C is a bilinear process such that y_t is a weak white noise (Granger and Andersen, 1978). In DGP B and DGP C the true state vector is wider than the one used in the Markov chain approximation ($L = 1$). In addition, we investigate the situation when $\alpha = .8$ and the data are generated by

$$\text{DGP D}: \quad y_t = .8 y_{t-1} + \varepsilon_t,$$

where ε_t is described above. To construct standard errors we use the HAC variance estimator of Newey and West (1987). The sample size T equals 30, and sometimes 60 and 120. The results are contained in Tables I and II.

We analyze the rejection frequencies obtained from $5,000$ experiments, for symmetric two-sided ("$\alpha \neq 0$") and lower-tail ("$\alpha < 0$") and upper-tail ("$\alpha > 0$") one-sides t-tests for the null $\alpha = 0$, and similarly for the null $\alpha = .8$. The MCB/LB critical values are computed from 500 replications of the MCB/LB algorithm. The first bootstrap observation is drawn randomly from the sample. For the MCB, we partition Ψ so that the observations are approximately evenly allocated across bins: $\Psi_k = \left[\psi_{k-1}, \psi_k\right)$, with $\psi_0 = y_{(1)}$, $\psi_k = \frac{1}{2}\left(y_{([kT/I])} + y_{([kT/I]+1)}\right)$, $k = 1, \ldots, I-1$, $\psi_I = y_{(T)} + 1$ (lines "MCB–I"). For the LB, we use the Epanechnikov resampling kernel (we also tried uniform and normal kernels as well, without noticeable changes in the results) with a regressor-specific bandwidth (a uniform bandwidth tends to worsen the results) set so that exactly n nearest neighbors are within the band (lines "LB–n"). The smoothing parameters are such that yield empirical sizes closest to nominal ones for two-sided alternatives. For reference, we also show rejection frequencies based on the asymptotic approximation (line "ASY").

Both MCB and LB yield comparable rejection probabilities that dominate asymptotic approximation for two-sided alternatives. The degree of precision falls, not dramatically though, when the true state vector differs from the one used. Interestingly, the LB performs slightly better for two-sided but slightly worse for one-sided tests than the MCB. Thus, in some situations a crude histogram may be preferable to a smooth resampling kernel. Note the persistent overrejection of upper-tail tests when the AR parameter is large, and this property does not seem to vanish with larger sample sizes. This may be caused by imprecise estimation of asymptotic variance during pivotization, the problem noticed by many researches (see, for instance, Kilian, 1999).

4. Application to US GNP

In this section we analyze the extended Nelson–Plosser US real per capita GNP yearly data from 1909 to 1988. The variable y_t is a first difference of natural logarithms of GNP and has 79 observations. Empirical studies of US GNP usually use threshold and smooth transition autoregressions, as well as Markov switching models (see Potter, 1995, for a review). Our purpose is not to search for the "true" or "best" model, but rather to demonstrate the application of the MCB in a simple model when the data are possibly generated by a complicated nonlinear mechanism. To this end, we choose a class of quadratic autoregressions.

From all autoregressions that are quadratic in the first three lags of y_t, the Akaike Information Criterion favors the model

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_3 y_{t-3} + \alpha_{23} y_{t-2} y_{t-3} + e_t.$$

The point estimates with the Newey–West standard errors are:

$$\hat{\alpha}_0 = \underset{(0.009)}{0.022}, \quad \hat{\alpha}_1 = \underset{(0.135)}{0.325}, \quad \hat{\alpha}_3 = \underset{(0.109)}{-0.278}, \quad \hat{\alpha}_{23} = \underset{(1.69)}{2.69}.$$

The t-statistics for $\hat{\alpha}_1$, $\hat{\alpha}_3$ and $\hat{\alpha}_{23}$ are 2.41, -2.55 and 1.59, respectively; the Wald statistic for joint significance of the three coefficients has the value of 14.48. The first two t-statistics and the Wald statistic exceed in absolute value even 1% asymptotic critical values. These results are likely to be regarded as an evidence of predictability of GNP by a person who relies on asymptotic inference with its tendency to overreject. However, the MCB with the parameter I equal to 7, 8, 9, gives the following 5% bootstrap critical values: 2.44, 2.50, 2.71 for $\hat{\alpha}_1$, 4.45, 4.50, 4.44 for $\hat{\alpha}_3$, 4.11, 4.20, 4.15 for $\hat{\alpha}_{23}$, and 40.47, 41.19, 40.49 for the Wald statistic. In contrast to asymptotics, the MCB inference says that there is little, if any, predictability of GNP within the chosen class of models.

References

Efron, B. (1979) "Bootstrap methods: another look at the jacknife" *Annals of Statistics* **7**, 1–26.

Granger, C.W.J., and A.P. Andersen (1978) *An introduction to bilinear time series models*, Vandenhoeck und Ruprecht: Göttingen, Germany.

Gregory, A.W. (1989) "A nonparametric test for autoregressive conditional heteroskedasticity: a Markov-chain approach" *Journal of Business and Economic Statistics* **7**, 107–114.

Hansen, B.E. (1999) "Non-parametric dependent data bootstrap for conditional moment models" University of Wisconsin–Madison working paper.

Horowitz, J.L. (2001) "Bootstrap methods for Markov processes" Northwestern University working paper.

Kilian, L. (1999) "Finite-sample properties of percentile and percentile-t bootstrap confidence intervals for impulse responses" *Review of Economics and Statistics* **81**, 652–660.

Newey, W.K., and K.D. West (1987) "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix" *Econometrica* **55**, 703–708.

Paparoditis, E., and D.N. Politis (2001a) "The local bootstrap for Markov processes" *Journal of Statistical Planning and Inference* **108**, forthcoming.

Paparoditis, E., and D.N. Politis (2001b) "A Markovian local resampling scheme for nonparametric estimators in time series analysis" *Econometric Theory* **17**, 540–566.

Potter, S. (1995) "A nonlinear approach to US GNP" *Journal of Applied Econometrics* **10**, 109–125.

Rajarshi, M.B. (1990) "Bootstrap in Markov-sequences based on estimates of transition density" *Annals of the Institute of Statistical Mathematics* **42**, 253–268.

Tauchen, G. (1986) "Finite state Markov-chain approximation to univariate and vector autoregressions" *Economics Letters* **20**, 177–181.

Table I. Actual rejection rates, nominal size 5%, DGPs A–C, $\alpha = 0$, $T = 30$

Inference method	DGP A			DGP B			DGP C		
	$\alpha \neq 0$	$\alpha < 0$	$\alpha > 0$	$\alpha \neq 0$	$\alpha < 0$	$\alpha > 0$	$\alpha \neq 0$	$\alpha < 0$	$\alpha > 0$
MCB–7	5.7	11.8	10.1	9.1	16.0	11.7	6.8	15.6	12.0
MCB–9	5.5	11.2	9.4	9.4	15.1	11.5	7.5	13.2	12.3
LB–4	5.8	13.0	11.2	8.3	15.1	12.2	6.5	14.2	13.7
LB–6	4.6	12.9	12.7	7.5	16.9	13.8	6.0	15.6	13.3
ASY	14.9	13.4	8.7	21.6	18.9	10.8	19.6	16.3	11.3

Table II. Actual rejection rates, nominal size 5%, DGP D, $\alpha = .8$, various T

Inference method	$T = 30$			$T = 60$			$T = 120$		
	$\alpha \neq .8$	$\alpha < .8$	$\alpha > .8$	$\alpha \neq .8$	$\alpha < .8$	$\alpha > .8$	$\alpha \neq .8$	$\alpha < .8$	$\alpha > .8$
		$I = 8$			$I = 10$			$I = 18$	
MCB–I	6.1	7.8	12.9	5.1	6.8	15.9	5.2	7.0	14.7
		$n = 5$			$n = 6$			$n = 7$	
LB–n	5.9	7.5	13.6	5.4	7.9	15.9	5.0	7.5	16.7
ASY	18.9	25.2	1.4	13.9	19.1	2.5	9.8	13.2	3.3