

Robustness of residual-based bootstrap to composition of serially correlated errors

by

Stanislav Anatolyev^{*}

New Economic School, Moscow

Abstract

In a simple autoregressive model with serially correlated errors, we evaluate size distortions resulting from the residual bootstrap when the Wold innovation is serially dependent and hence is expected to contaminate the inference. Small distortions caused by the presence of strong conditional heteroskedasticity or other nonlinearities can be partly removed further by using the wild bootstrap.

Key words: Residual bootstrap, wild bootstrap, serial correlation, rejection rate

^{*}Address: Nakhimovsky Prospekt, 47, office 1721, Moscow 117418 Russia. E-mail: sanatoly@nes.ru.

I thank Alexander Malyi for research assistance.

1 Introduction

For models with serially correlated errors, the residual bootstrap (Bose 1988, 1990) presumes an ideal moving average structure of the error, that the Wold innovations are serially independent and hence may be harmlessly reshuffled during resampling. Such error structure may not hold in reality, and the primary purpose of this note is to explore how critical this condition is, and what happens when the error does not have that ideal property. We evaluate, via simulations, size distortions resulting from the use of residual bootstrap resampling when the error composition is such that the Wold innovation is serially dependent.

The critical feature of the situation we concentrate on is the presence of serial correlation of finite order, but without further knowledge of the structure of the error term. Bose (1990) studied residual bootstrap in moving average models with ideal structure; besides, he was interested in inference about moving average coefficients. We are instead concerned with the inference about coefficients in the conditional mean while the serial correlation in the error is treated as a nuisance feature. To isolate distortions caused *only* by the structure of the error term, we use a very simple linear model without exogenous variables and with the slope parameter equal to zero. We tune the parameters of the Data Generating Process (DGP) so that the regression error is linear or nonlinear, conditionally homo- or heteroskedastic, with innovations that are serially independent or only serially uncorrelated but dependent. As a measure of performance, we use closeness of actual rejection rates to nominal sizes for testing the null that the slope parameter is zero. The parameters are estimated by OLS, and the corresponding t -ratio is bootstrapped.

Along with the residual bootstrap (Bose 1988), we also consider the wild bootstrap (Wu 1986). While the residual bootstrap is ideal when the Wold innovation in the error is a serially independent sequence, the wild bootstrap is expected to help in the case of a conditionally heteroskedastic error term. The simulation evidence shows that the residual bootstrap performs well even in situations where the non-IID structure of Wold innovations is expected to contaminate the inference. Small distortions caused by the

presence of conditional heteroskedasticity or other nonlinearities are partly removed by the wild bootstrap.

The rest of the paper is organized as follows. Section 2 presents the model and test statistic, section 3 – four DGPs. Section 4 describes the bootstrap algorithms. In section 5 simulation results are reported and discussed. Section 6 concludes.

2 Model and test statistic

The working model is linear with the error that is first-order serially correlated:

$$y_{t+2} = \alpha + \beta y_t + e_{t+2}, \quad E_t[e_{t+2}] = 0, \quad (1)$$

where $E_t[\cdot] \equiv E[\cdot | y_t, y_{t-1}, \dots]$. The true values of α and β are set to zero in all DGPs. The zero value of β corresponds to testing the null of no predictive ability and allows us not to be distracted by an autoregression bias which is often blamed for unsatisfactory bootstrap performance (see Kilian 1999). Note that the model is silent about the composition of the error term e_{t+2} . In particular, while it is representable in the MA(1) form with serially uncorrelated Wold innovations according to the Wold decomposition theorem, the innovations may not be serially independent.

The estimator we concentrate on is the OLS estimator of β :

$$\hat{\beta} = \frac{\sum_{t=1}^{T-2} y_{t+2}(y_t - \bar{y})}{\sum_{t=1}^{T-2} (y_t - \bar{y})^2}, \quad (2)$$

where $\bar{y} = \frac{1}{T-2} \sum_{t=1}^{T-2} y_t$, and T is the sample size. The OLS estimator of α is $\hat{\alpha} = \frac{1}{T-2} \sum_{t=1}^{T-2} y_{t+2} - \hat{\beta} \bar{y}$. The residuals are computed as $\hat{e}_1 = y_1 - \hat{\alpha} - \hat{\beta} \bar{y}$, $\hat{e}_2 = y_2 - \hat{\alpha} - \hat{\beta} \bar{y}$, $\hat{e}_t = y_t - \hat{\alpha} - \hat{\beta} y_{t-2}$, $t = 3, \dots, T$. We use the following simple variance estimator: let $\vec{y}_t = (1 \ y_t)'$, then

$$\hat{V} = \left(\sum_{t=1}^{T-2} \vec{y}_t \vec{y}_t' \right)^{-1} \left(\sum_{t=1}^{T-2} \vec{y}_t \vec{y}_t' (\hat{e}_{t+2})^2 \right) + \sum_{t=1}^{T-3} \left(\vec{y}_{t+1} \vec{y}_t' + \vec{y}_t \vec{y}_{t+1}' \right) \hat{e}_{t+2} \hat{e}_{t+3} \left(\sum_{t=1}^{T-2} \vec{y}_t \vec{y}_t' \right)^{-1}, \quad (3)$$

(we omit scalar factors like $T - 2$ since they are immaterial for the bootstrap). This is the familiar Hansen and Hodrick (1980) estimator that takes advantage of the structure

of error autocorrelation: the middle term in (3) is a (scaled) estimate of the variance and first-order autocovariances of $\vec{y}_t e_{t+2}$. Because the Hansen–Hodrick estimator is not guaranteed to be positive definite, we modify \widehat{V} by excluding the covariance terms in the middle term in (3) whenever the (2, 2) component $\widehat{V}_{2,2}$ is negative. The t -statistic is $t_\beta = \widehat{\beta} / \sqrt{\widehat{V}_{2,2}}$.

3 Data generating processes

We consider the following four types of structure of the error e_t in the DGP: linear moving average with IID innovations; conditionally homoskedastic with uncorrelated but not independent innovations; conditionally heteroskedastic with ARCH-type form of heteroskedasticity, and nonlinear moving average with IID fundamental shocks. In all DGPs, the parameters are set so that e_t has the same variance and autocovariance structure: $E[e_t^2] = 1 + \theta^2$, $E[e_{t+1}e_t] = -\theta$, $E[e_{t+j}e_t] = 0$ for $|j| > 1$, where $|\theta| < 1$, so that it is representable as $e_{t+2} = \varepsilon_{t+2} - \theta\varepsilon_{t+1}$, where ε_t is (weak) white noise. In all DGPs, the errors are leptokurtic reflecting one of specifics of financial data; the presence of error autocorrelation is also characteristic of many financial data models (see Campbell, Lo and MacKinlay 1997).

The simplest structure of the error occurs when it is a moving average with independent innovations. We will call this *DGP IID*:

$$e_{t+2} = w_{t+2} - \theta w_{t+1}, \quad w_{t+2} \sim IID \sqrt{.6} \cdot t_{(5)}. \quad (4)$$

The Wold innovation ε_{t+2} equals w_{t+2} , a serially independent sequence. In another error structure, the innovations are not independent. We will call this *DGP UC*:

$$\begin{aligned} e_{t+2} &= u_{t+2} - \text{sgn}(\theta)u_{t+1} + v_{t+2}, \\ u_{t+2} &\sim IID \sqrt{.6|\theta|} \cdot t_{(5)}, \quad v_{t+2} \sim IID \sqrt{.6}(1 - |\theta|) \cdot t_{(5)}, \end{aligned} \quad (5)$$

where u_{t+2} and v_{t+2} are independent. The principal difference between error structures in (4) and (5) is that the Wold innovation in (4) is an IID sequence (a structure that is ideal for the residual bootstrap), while in (5) the innovation is a serially uncorrelated,

but not independent, sequence (a structure that may potentially invalidate the residual bootstrap). Indeed, the Wold innovation equals

$$\varepsilon_{t+2} = \sum_{i=0}^{\infty} \theta^i e_{t+2-i} = u_{t+2} + (\theta - \text{sgn}(\theta)) \sum_{i=0}^{\infty} \theta^i u_{t+1-i} + \sum_{i=0}^{\infty} \theta^i v_{t+2-i}$$

and is clearly serially dependent (serial independence would follow from uncorrelatedness only if the two disturbances u_t and v_t were normally distributed). Student's distribution in (4) and (5) leads to leptokurticity of innovations.

Next, we explore a conditionally heteroskedastic structure of the error calling this *DGP HS*:

$$e_{t+2} = \varrho_{t+2} - \theta \varrho_{t+1}, \quad \varrho_{t+2} = \zeta_{t+2} \sqrt{\omega_t}, \quad \zeta_{t+2} \sim IID \mathcal{N}(0, 1), \quad (6)$$

where the auxiliary process ω_t indexing conditional heteroskedasticity is specified in the ARCH(1) spirit: $\omega_t = 1 - \alpha_\omega(1 + \theta^2) + \alpha_\omega e_t^2$, where $0 < \alpha_\omega < 1$. The conditional autocovariance structure of e_{t+2} is: $E_t[e_{t+2}^2] = \omega_t + \theta^2 \omega_{t-1}$, $E_t[e_{t+2} e_{t+1}] = -\theta \omega_{t-1}$. We put $\alpha_\omega = 0.3$ so that the fourth moment of e_{t+2} is finite for all values of θ such that $|\theta| < 1$. The Wold innovation $\varepsilon_{t+2} = \varrho_{t+2}$ is clearly serially dependent. The ARCH structure induces leptokurticity even though the conditional distribution is normal.

Finally, the MA(1) error can be generated nonlinearly from IID random sequence. We call this *DGP NL*:

$$e_{t+2} = \varsigma_{t+2} \varsigma_{t+1} - \theta \varsigma_{t+1} \varsigma_t, \quad \varsigma_{t+2} \sim IID \mathcal{N}(0, 1). \quad (7)$$

The Wold innovation $\varepsilon_{t+2} = \varsigma_{t+2} \varsigma_{t+1}$ is again serially dependent. The innovations are leptokurtic because $E[(\varsigma_{t+1} \varsigma_t)^4] = 9 > 3$.

To summarize, the Wold innovation is IID under the DGP (4), but it is not under the DGPs (5), (6) or (7).

4 Bootstrap resampling

In the *residual bootstrap* one resamples the Wold innovation in the error treated as an IID process (Bose 1990, Kreiss and Franke 1992). After the residuals \hat{e}_t , $t = 1, \dots, T$ are

computed, we restore estimates of the Wold innovations $\hat{\varepsilon}_t$, $t = 1, \dots, T$ in the following way. We compute an estimate of θ by the method of moments imposing restrictions on a value of the correlation coefficient (as the population correlation coefficient is bounded by $\frac{1}{2}$ in absolute value), which allows us to compute the estimate in a closed form: $\hat{\theta} = -2\hat{\rho}/(1 + (1 - 4\hat{\rho}^2)^{1/2})$, where $\hat{\rho} = \min(.499, \max(-.499, (\sum_{t=3}^{T-1} \hat{\varepsilon}_t \hat{\varepsilon}_{t+1})/(\sum_{t=3}^{T-1} \hat{\varepsilon}_t^2))$. Then we calculate innovations: $\hat{\varepsilon}_t = \sum_{i=0}^{t-1} \hat{\theta}^i \hat{\varepsilon}_{t-i}$, $t = 1, \dots, T$. We then resample $\hat{\varepsilon}_t$ from the original sample randomly, uniformly over t , with replacement. Having obtained a bootstrap sample ε_t^* , $t = 1, \dots, T$, we generate the e^* and y^* series recursively as $e_1^* = \varepsilon_1^*$, $y_1^* = \hat{\alpha} + \hat{\beta}\bar{y} + e_1^*$, $e_2^* = \varepsilon_2^* - \hat{\theta}\varepsilon_1^*$, $y_2^* = \hat{\alpha} + \hat{\beta}\bar{y} + e_2^*$, $e_t^* = \varepsilon_t^* - \hat{\theta}\varepsilon_{t-1}^*$, $y_t^* = \hat{\alpha} + \hat{\beta}y_{t-2}^* + e_t^*$, $t = 3, \dots, T$. Using the bootstrap sample we obtain the bootstrap OLS estimator $\hat{\beta}^*$, bootstrap variance estimate \hat{V}^* by (2) and (3) evaluated at the bootstrap sample, and bootstrap t -statistic $t_\beta^* = (\hat{\beta}^* - \hat{\beta})/\sqrt{\hat{V}_{2,2}^*}$.

The *wild bootstrap* proposed by Wu (1986) helps to preserve the pattern of conditional heteroskedasticity in bootstrap samples. In the context of an autoregression it was described in Kreiss (1997) and applied, for instance, in Hafner and Herwartz (2000). We adapt the algorithm to our MA(1) setting. The construction of a bootstrap sample is similar to that for the residual bootstrap, but instead of resampling bootstrap innovations from the set of estimated innovations $\hat{\varepsilon}_t$, $t = 1, \dots, T$, we obtain them by multiplying the latter by an IID zero mean sequence η_t , $t = 1, \dots, T$ having properties $E[\eta_t^2] = E[\eta_t^3] = 1$, i.e. $\varepsilon_t^+ = \eta_t \hat{\varepsilon}_t$, $t = 1, \dots, T$. Then we set $e_1^+ = \varepsilon_1^+$, $e_2^+ = \varepsilon_2^+ - \hat{\theta}\varepsilon_1^+$, $e_t^+ = \varepsilon_t^+ - \hat{\theta}\varepsilon_{t-1}^+$, $t = 3, \dots, T$. A bootstrap sample is generated recursively: $y_1^+ = \hat{\alpha} + \hat{\beta}\bar{y} + e_1^+$, $y_2^+ = \hat{\alpha} + \hat{\beta}\bar{y} + e_2^+$, $y_t^+ = \hat{\alpha} + \hat{\beta}y_{t-2}^+ + e_t^+$, $t = 3, \dots, T$. From the bootstrap sample we obtain the bootstrap OLS estimator, bootstrap variance estimate and bootstrap t -statistic as in the residual bootstrap. In our experiment, we use the following probability distribution for η_t : let (η_{1t}, η_{2t}) be standard bivariate normal, then $\eta_t = \eta_{1t}/\sqrt{2} + (\eta_{2t}^2 - 1)/2$ (Mammen 1993).

5 Simulation results

We evaluate rejection rates for 5% size tests on the basis of 10,000 simulations (the results for 1% and 10% size tests exhibit similar patterns). Thus, the estimates will have a standard error of approximately $\sqrt{5\% \cdot (100\% - 5\%)/10,000} \approx 0.22\%$. In a single simulation loop, a time series for y_t of $T + 1,000$ observations with zero starting values is generated, first 1,000 observations are discarded, and 1,000 repetitions are used to form a bootstrap distribution and read off bootstrap critical values. Table 1 reports actual rejection frequencies for symmetric two-sided alternatives, i.e. $\Pr\{|t_\beta| > q_{5\%}^*\}$, where $q_{5\%}^*$ is an appropriate critical value corresponding to the nominal size of 5%. For reference, we also give actual rejection rates for the asymptotic approximation. We set the sample length to 20, 40, 80, and MA coefficient to 0, ∓ 0.3 , ∓ 0.6 , ∓ 0.9 . This allows us to study the impact of a sample length and strength of serial correlation. Each row of Table 1 contains actual rejection frequencies for a particular sample size T and serial correlation parameter θ . Columns IID, UC, HS and NL correspond to DGPs described in section 3.

One can notice immediately that bootstrap rejection rates are much closer to nominal sizes than asymptotic ones, especially for smaller sample lengths, and than those frequently encountered in other bootstrap simulation studies. The latter fact is a consequence of linearity and absence of autoregressive persistence. Generally, size distortions tend to rise as the degree of serial correlation (value of $|\theta|$) increases, more strongly for positive serial correlation ($\theta < 0$) than for negative ($\theta > 0$).

Consider the panel “Residual bootstrap”. The DGP IID has an ideal error structure for the residual bootstrap, hence the corresponding size distortions can be considered as lower bounds for distortions for other DGPs. However, the results for the DGP UC are nearly identical to those for the DGP IID in spite of serial dependence of Wold innovations. This implies a perhaps surprising fact: serial uncorrelatedness seems to be a guarantee against big distortions in the residual bootstrap. Practically no distortions are observed for both DGPs when $\theta = 0$ even for the smallest sample size. For the DGP HS and DGP NL they are but slightly worse: the actual rejection rates deviate from the nominal 5% by no more

than 3%.

Consider now the panel “Wild bootstrap”. The “ideal” numbers for the DGP IID and DGP UC are worsened by use of the wild bootstrap, although very slightly. Interestingly, the evidence for the DGP HS where the wild bootstrap is hoped to correct for distortions caused by conditional heteroskedasticity is mixed. The wild bootstrap tends to do so only for larger sample sizes, but can actually worsen the situation for smaller ones. It appears that the wild bootstrap does a better job for the DGP NL where the dependence structure in the error is less clear.

6 Conclusion

To summarize, the residual bootstrap performs well even in situations where the non-IID structure of the error may be expected to contaminate the inference much more. The distortions are very small, especially when compared to those arising from nonlinearities in the conditional mean or its high persistence. Even these small distortions caused by the presence of conditional heteroskedasticity or other nonlinearities can be partly removed by using the wild bootstrap, at least for non-tiny samples.

References

- Bose, A. (1988) Edgeworth correction by bootstrap in autoregressions. *Annals of Statistics*, 16, 1709–1722.
- Bose, A. (1990) Bootstrap in moving average models. *Annals of the Institute of Statistical Mathematics*, 42, 753–768.
- Campbell, J., Lo, A., and MacKinlay, A.C. (1997) *The Econometrics of Financial Markets*. Princeton University Press, Princeton.
- Hafner, C.M. and Herwartz, H. (2000) Testing for linear autoregressive dynamics under heteroskedasticity. *Econometrics Journal*, 3, 177–197.
- Hansen, L.P and Hodrick, R.J. (1980) Forward exchange rates as optimal predictors of future spot rates: an econometric analysis. *Journal of Political Economy*, 88, 829–853.

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.

Kreiss, J.-P. (1997) Asymptotic properties of residual bootstrap for autoregressions. Institute for Mathematical Stochastics, Technical University of Braunschweig. Available at <http://www.math.nat.tu-bs.de/stochastik/publ.htm>.

Kreiss, J.-P. and Franke, J. (1992) Bootstrapping stationary autoregressive moving-average models. *Journal of Time Series Analysis*, 13, 297–317.

Mammen, E. (1993) Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics*, 21, 255–285.

Wu, J. (1986) Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics*, 14, 1261–1343.

Table 1. Actual rejection rates for asymptotic and bootstrap inferences

		Asymptotic approximation				Residual bootstrap				Wild bootstrap			
T	θ	IID	UC	HS	NL	IID	UC	HS	NL	IID	UC	HS	NL
20	-0.9	19.9	20.8	20.5	20.1	8.2	8.6	8.9	7.5	8.8	8.6	9.0	7.9
	-0.6	19.3	18.3	18.9	21.0	8.0	7.4	8.2	8.9	8.0	8.7	8.5	9.0
	-0.3	15.2	14.6	15.9	16.6	6.2	6.1	6.8	6.6	7.0	7.3	6.6	7.0
	0	13.3	13.7	14.2	13.2	4.9	5.0	5.4	4.8	6.0	6.0	6.4	4.7
	0.3	13.5	13.3	14.5	15.0	4.9	5.0	5.8	5.6	6.3	6.0	6.4	5.9
	0.6	13.9	14.1	15.5	16.4	5.6	5.1	7.1	7.1	5.9	6.2	6.4	6.3
	0.9	14.4	14.5	16.2	17.2	5.3	5.4	6.5	6.8	6.1	6.0	7.6	6.0
40	-0.9	12.4	12.4	14.3	15.4	6.6	6.4	7.6	7.9	6.1	5.7	6.7	6.6
	-0.6	12.0	12.5	13.3	15.9	6.5	6.0	7.2	8.5	6.2	6.0	6.7	7.3
	-0.3	10.5	10.0	11.6	13.8	5.9	5.6	6.2	7.3	6.3	6.1	6.4	6.2
	0	9.3	9.6	10.4	10.9	4.8	5.0	5.7	5.4	5.9	5.7	6.1	5.2
	0.3	9.6	10.2	10.6	11.9	5.7	4.9	6.2	6.8	5.3	5.4	5.8	5.2
	0.6	9.9	10.4	11.4	13.2	5.2	5.3	6.3	7.2	5.3	5.0	6.3	5.8
	0.9	10.4	10.0	12.1	12.5	5.4	5.0	7.5	6.8	5.1	4.9	6.6	5.0
80	-0.9	9.4	9.0	10.5	11.4	5.0	5.8	7.0	7.0	5.0	5.4	6.0	4.9
	-0.6	8.6	9.1	9.9	11.8	5.6	5.4	5.9	8.0	5.7	5.3	5.6	5.2
	-0.3	8.3	7.9	8.8	10.8	5.1	5.6	6.0	7.0	4.7	5.1	5.7	5.6
	0	7.4	7.0	8.1	9.2	5.1	5.2	5.8	6.3	5.2	5.1	5.8	4.7
	0.3	7.7	7.6	8.2	10.0	5.2	5.0	5.8	7.2	4.9	4.9	5.7	5.4
	0.6	7.7	7.7	9.0	10.1	5.1	5.1	5.9	6.6	4.4	4.7	5.7	5.5
	0.9	7.6	8.1	9.6	10.0	5.3	5.0	6.4	6.6	4.7	4.7	5.8	4.5

Notes: Each row contains actual rejection frequencies for a particular sample size T and serial correlation parameter θ , for a nominal size of 5%. Columns IID, UC, HS and NL correspond to DGPs described in section 3