Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Journal of Econometrics 133 (2006) 411–419
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Editors' Introduction

Resampling methods in econometrics

Resampling is used in statistical analysis and econometrics for two main purposes. The first one is to improve usual asymptotic approximations in situations where such approximations are in principle valid but are not reliable in finite samples. The second one is to allow inference in cases where an asymptotic distribution is not available or depends on unknown nuisance parameters which are difficult to eliminate. Without the use of resampling methods, a different set of critical values would have to be computed for each possible value of these nuisance parameters. In certain cases, resampling can even lead to exact (randomized) tests and confidence sets (Monte Carlo tests), whose level is fully controlled in finite samples, while in other ones considerable improvements can be achieved with respect to usual asymptotic approximations (bootstrap), a feature which is often demonstrated by showing "bootstrap refinements".

Resampling methods are typically numerically intensive, thus making them more easily applicable in an environment with easier access to computing power. This heavy use of the computer also explains in part the rapid explosion of academic research in recent years, most notably for increasingly complex nonlinear and/or dynamic models. Yet, despite its tight link with computing technology, resampling has been around for much longer than the computer. The first use of the bootstrap in economics that we are aware of is in a paper by Cowles (1933) in his analysis of the performance of stock market forecasters where he used drawing cards to generate his random samples from observed data. It is also interesting to note that Monte Carlo tests go back to Dwass (1957) and Barnard (1963), while bootstrapping was subsequently introduced by Efron (1979).

This volume brings together many papers that were presented at a conference held under the auspices of the Centre de Recherche et Développement en Économique (CRDE) which has now become the Centre Interuniversitaire de Recherche en Économie Quantitative (CIREQ) on October 13–14, 2001. The aim of the conference was to bring together applied and theoretical econometricians and statisticians who work in this area. The conference program included 24 presentations, including the 16 papers that appear in this issue. Many of the other contributions have been or will be published in other journals of high quality (or even other issues of the journal).

doi:10.1016/j.jeconom.2005.06.001

The collection of papers in this issue are organized around four themes:

(1) General issues on the reliability and power of resampling procedures in finite samples (Davidson and MacKinnon; Dufour)
(2) Inference in nonlinear models and GMM estimation (Jouneau-Sion and Torrès; Luger; Inoue and Shintani; Hong and Scaillet; Yatchew and Härdle)
(3) Unit roots, cointegration and long memory (Andrews, Lieberman, and Marmer; Chang, Park, and Song; Davidson; Park; Parker, Paparoditis, and Politis)
(4) Specification tests (Corradi and Swanson; Hidalgo and Kreiss; Horowitz, Lobato, Nankervis, and Savin; Li and Tcakz)

The ordering of the papers follows thematic considerations and does not in any way represent an assessment of the relative merits of the contributions made. We shall now succinctly review the contents of the special issue.

1. General issues on the reliability and power of resampling procedures in finite samples

The first two papers (Davidson and MacKinnon; Dufour) deal with general issues associated with the use of resampling methods in view of getting more reliable tests.

Davidson and MacKinnon study the problem of assessing the difference between a bootstrap test and the "ideal" test based on the "true distribution" of a test statistic. For that purpose, they introduce the concept of *bootstrap discrepancy*, which measures the difference in rejection probabilities between a bootstrap test and the corresponding "level-adjusted" asymptotic test. Under fairly general conditions, they show that the bootstrap discrepancy converges to zero at the same rate under the null hypothesis and under local alternatives (following a Pitman drift). If the test statistic is not an exact pivot, critical values depend on which data-generating process (DGP) is used to determine the null distribution. They propose using the DGP which minimizes the bootstrap discrepancy. They also show that, under an asymptotic independence condition, the power of both bootstrap and asymptotic tests can be estimated cheaply by simulation. Finally, the results are illustrated by Monte Carlo simulations using a logit model.

The other paper considers the problem of building exact tests through resampling methods. For that purpose, the paper reconsiders the technique of Monte Carlo (MC) tests (Dwass, 1957; Barnard, 1963), which allows perfect size control provided the test statistic can be simulated under the null hypothesis (H_0). This approach is, however, restricted to test statistics whose distribution is either continuous or discrete in specific ways. More importantly, the exactness result does not hold when the null distribution depends on nuisance parameters, which appears quite restrictive in econometric models. The paper extends this approach in two ways: first, by allowing for MC tests based on exchangeable statistics whose form is completely arbitrary; second, by generalizing the method to statistics whose null distributions involve nuisance parameters, which yields a technique called *maximized Monte Carlo*

(MMC) tests. This technique is applicable as soon as the test can be simulated under H_0 given a finite number of parameters, and thus covers in principle all parametric models as well as nonparametric (or semiparametric) models for which distribution-free statistics are available (such as sign and rank statistics). The number of replications required may be quite small (e.g., as low as 19 for a test with level 0.05) and the MMC maximization required may be achieved by any algorithm applicable to nondifferentiable functions (such as simulated annealing). Simplified asymptotically justified versions of the MMC method are also proposed and it is shown that they provide a simple way of dealing with nonstandard asymptotics (e.g., unit root asymptotics).

2. Inference in nonlinear models and GMM estimation

The next five papers study the application of resampling methods to inference on nonlinear models and to GMM estimation. The first two (Jouneau-Sion and Torrès; Luger) propose exact methods, the next two consider asymptotic resampling methods (Inoue and Shintani; Hong and Scaillet), while the last one (Yatchew and Härdle) considers a nonparametric regression setup.

Jouneau-Sion and Torrès apply the MMC approach to dichotomous choice models with a latent variable, in order to derive exact tests and confidence regions. In view of the fact that the MMC method may require a difficult numerical optimization involving a nondifferentiable objective function, they show that the latter can be reduced in such models to a mixed integer programming problem (MIP), a class of problems recently introduced in the operations research literature. Powerful algorithms are now available to solve such optimizations, such as the branch-and-bound algorithm. The authors show a numerical illustration of their method on a probit model. They also show that their method can be extended to multinomial logit models, the Poisson counting models, and disequilibrium models.

The article by Luger proposes exact distribution-free tests for a possibly nonlinear regression model with exchangeable disturbances, against one (or several) nonnested alternatives. The null hypothesis sets the value of the parameters that appear in the regression, which is tested against a (possibly) nonnested nonlinear alternative, using the J statistic proposed by Davidson and MacKinnon (1981). For such statistics, it is well known that usual asymptotic approximations are quite unreliable even under parametric distributional assumptions. Luger succeeds in getting exact procedures by combining two techniques: first, a distribution-free procedure is obtained by considering the permutational distribution of the J statistics; second, in view of the complicated structure of the null distribution of the test, the latter is implemented as a MC test (taking into account the discrete nature of the distribution). In this way, the severe size distortions which typically occur in such contexts are completely eliminated. A simulation experiment confirms the reliability of the new test procedure, and, overall, its power is practically identical to that of the size-corrected

J test. The proposed test procedure is illustrated with an empirical example on inflation dynamics.

Inoue and Shintani study the performance of an overlapping block bootstrap procedure applied to GMM-based t and J statistics in the context of an overidentified linear model (or an IV linear regression), when the autocorrelation structure of the moment functions is unknown. When moment functions are uncorrelated after finite lags, Hall and Horowitz (1996) showed that errors in the rejection probabilities of the bootstrap tests are $o(n^{-1})$, where n is the sample size. However, this rate does not hold in general with the HAC covariance matrix estimator, since it converges at a nonparametric rate. By incorporating the HAC covariance matrix estimator in the Edgeworth expansion of the distribution, the authors show that the bootstrap does indeed provide asymptotic refinements when the characteristic exponent of the kernel function is greater than two. They also present a small MC experiment illustrating the improvements associated with bootstrapping as well as an application to the monetary policy reaction function in the United States.

The paper by Hong and Scaillet suggests an implementation of subsampling in nonlinear models which is computationally attractive since it avoids the need for numerical optimization. The thrust of the algorithm is to avoid numerical optimization by computing the scores on the subsampled data. This is an idea that has already been used in the bootstrap literature; see, for example, Davidson and MacKinnon (1999). The only requirement for the validity of this method is that the limiting distribution of the estimator admits a linear representation asymptotically. The authors also demonstrate that their method applies even when the rate of convergence of the estimator is unknown and needs to be estimated as in Bertail et al. (1999). Simulation results document the good finite sample performance of the method even when usual subsampling is available and its great potential in reducing computation time.

The paper by Yatchew and Härdle discusses nonparametric estimation of functions under shape constraints. This paper suggests an estimator that minimizes a quadratic form subject to a smoothness condition. The authors replace this problem by an equivalent one that reduces to a least-square projection onto representators. The authors then argue that shape constraints (such as monotonicity and convexity) can be easily imposed in this framework. Finally, they propose tests for the validity of these restrictions. Since asymptotic inference appears to be very unreliable, the authors suggest using the bootstrap to improve inference in finite samples. The proposed method is applied to the call function for European options. Economic theory entails constraints on the shape of this function (monotonicity and convexity), but unless one makes strong assumptions, no precise functional form emerges (as in the Black and Scholes (1973) pricing formula for example). Aït-Sahalia and Duarte (2003) have considered this problem with local polynomial estimators. Interestingly, the authors find it useful to impose "tail" constraints restricting the state-price density (which is proportional to the second derivative of the call function) to be unimodal. The procedure is finally illustrated by estimating the relation between option prices and strike prices on German data.

3. Unit roots, cointegration and long memory

The next five papers study applications of resampling techniques to linear models with highly dependent or nonstationary dynamics. The first paper (Parker, Paparoditis and Politis) proposes a bootstrap-based unit root test. The next two (Park; Andrews, Lieberman, and Marmer) consider inference in long memory and near-unit root processes. Finally, the last two papers (Chang, Park, and Song; Davidson) consider cointegration among nonstationary variables.

The paper by Parker, Paparoditis, and Politis develops a residual-based bootstrap test of the null hypothesis of a unit root. The authors rely on the stationary bootstrap (Politis and Romano, 1994) to resample the residuals and account for serial dependence in them. The resulting bootstrap sample is generated by imposing the null hypothesis of a unit root by summing the resampled residuals, and the validity of the procedure is established under general conditions. The authors describe in detail the case where the autoregressive parameter is estimated by ordinary least squares and the impact of assumptions on the deterministic component of the process. The new test is shown to be consistent, and its power against standard local alternatives (shrinking towards the unit root at rate n^{-1}) is a function of an Ornstein–Uhlenbeck process. Further, this test dominates a procedure based on differences suggested by Swensen (2003) in terms of local power. Simulation evidence confirms this theoretical result but also finds that the bootstrap test based on differences suffers less from over-rejection in the usual problematic case of a large negative MA root (Schwert, 1989).

Bootstrapping unit root processes is a very difficult problem. Basawa et al. (1991) have shown that the bootstrap typically fails in this case if it is based on the estimated autoregressive parameter. However, the paper by Park develops a bootstrap theory for autoregressive models with roots close to 1. The theoretical framework rests on roots that are local to unity but converge towards 1 at a rate slower than the usual rate of n^{-1} used for nonstationary processes. Park defines such processes as weakly integrated. Park establishes that, contrary to pure unit root processes, the bootstrap based on the estimated autoregressive model is generally consistent for these weakly integrated processes. In addition, for statistics that are pivotal, the author shows that the bootstrap will yield asymptotic refinements with the magnitude of these positively related to the rate at which the autoregressive root converges to 1.

The paper by Andrews, Lieberman, and Marmer analyzes the behavior of the bootstrap for data with long memory. A series is called long memory (or fractionally integrated) if its autocorrelation function decays at a hyperbolic rate for large lag values, i.e. $\rho_h \sim c_1 h^{2d-1}$ as $j \to \infty$ for some constant c_1. The parameter d is called the memory parameter (or order of integration) since it controls the rate of decay of the autocorrelation function for large lags. Processes with $d > \frac{1}{2}$ are nonstationary. The results in this paper are restricted to Gaussian (stationary) processes with $d \in (0, \frac{1}{2})$. For this case, the authors develop valid higher-order expansions which allow them to show that the parametric bootstrap delivers asymptotic refinements compared to confidence intervals built using the usual asymptotic normal distribution. Since maximum likelihood is very difficult to compute in such a setup (see Sowell, 1992, for example), two commonly used approximate estimators are considered, the "plug-in" likelihood with the unknown mean replaced by the sample mean and the "plug-in" Whittle likelihood. The paper finds the interesting result that the error in coverage rates of one-sided confidence intervals is almost the same as the one for i.i.d. data shown by Hall (1992). The only difference is a term of order $\ln(n)$ which the authors conjecture is due to their method of proof. This result is however not true for two-sided confidence intervals where the improvement is smaller than the one for i.i.d. data.

The paper by Chang, Park, and Song analyzes the difficult problem of bootstrapping cointegrating regressions. Cointegration is imposed in the resampling procedure by bootstrapping the pairs $(\hat{u}_t, v_t)'$ where $\hat{u}_t$ are the residuals from the cointegrating regressions and v_t are the first differences of the regressors as suggested by Li and Maddala (1995). This paper uses a sieve bootstrap on this vector which consists of fitting a $VAR(p)$ process with the order p of the process increasing with sample size. Thus, asymptotically, this approach can capture very general serial dependence. In cointegrating regressions, it is known that standard OLS has an asymptotic distribution which depends on nuisance parameters stemming from the presence of endogeneity and serial correlation (see Park and Phillips, 1998). Inference using this estimator is therefore very difficult. This paper shows that the bootstrap as implemented above is valid in this case, which makes bias-correction and inference using OLS possible. As is well known also, OLS is not the efficient estimator in this case. The paper shows that the bootstrap is valid for the efficient estimators of Saikkonen (1991) and Stock and Watson (1993) based on augmenting the regression with past, present, and future first differences of the regressors (v_t). Simulations demonstrate that the bootstrap is able to control size for statistics based on either the OLS or the Saikkonen estimator.

Finally, the paper by Davidson considers various bootstrap tests of cointegration among fractionally integrated processes. This paper considers fractional cointegration among nonstationary processes, i.e. $\frac{1}{2} < d < \frac{3}{2}$. The resampling is parametric in the sense that it requires the specification of the short-run dynamics via an error correction model. In addition to the usual bootstrap, the author considers the double bootstrap and the fast double bootstrap of Davidson and MacKinnon (2000). The paper investigates residual-based tests for the null of no cointegration and the null of cointegration. Three statistics are considered: the Durbin–Watson statistic of the residuals, the F statistic of the first-stage regression, and the Kwiatkowski et al. (1992) statistic. None of these statistics is asymptotically pivotal. Limited simulation experiments show that all three statistics lead to tests that have good size in almost all cases considered. Surprisingly at first, the double bootstrap does not improve uniformly relative to the performance of the regular bootstrap. The author attributes this behavior to misspecification of the model used to generate the bootstrap data: the simple bootstrap is less affected by possible misspecification than the other methods. The author reassesses the finding of cointegration among consumption, labor income, and net wealth documented by Lettau and Ludvigson (2001). The residuals from such a regression, called the consumption–wealth ratio and denoted by *cay*, has been used to forecast future returns on aggregate stock markets. Davidson concludes that the three variables share a fractional cointegrating relation with each variable integrated of order greater than 1 and a linear combination of them being integrated of order less than 1 (but still nonstationary).

4. Specification tests

The final four papers (Corradi and Swanson; Hidalgo and Kreiss; Horowitz, Lobato, Nankervis, and Savin; Li and Tcakz) propose bootstrap procedures for testing goodness of fit or, more generally, the specification of a model.

The paper by Corradi and Swanson study bootstrap versions of two tests of a conditional distribution when the conditioning information is incomplete. This feature is called "dynamic misspecification" by the authors, and it induces dependence among the data transforms (such as conditional probability integral transforms) used to construct the test statistics. The authors study two tests: a version of the DGT test of Diebold et al. (1998) based on fitted conditional empirical distributions, and the CK test of Andrews (1997) which considers marginal empirical distribution functions. Under strict stationarity and some additional regularity conditions on the shape of the conditional distribution functions, Corradi and Swanson first establish the asymptotic distribution of the test statistics involved, which can be characterized as Gaussian processes with covariance kernels that reflect dynamic misspecification and parameter estimation error. Since these kernels involve nuisance parameters, they propose a block-bootstrap approach and show that it is valid to first order. Critical values are based on an extension of the empirical process version of the block bootstrap to the case of non vanishing parameter estimation error. Finally, the authors present the results of a MC experiment which indicate that the finite-sample properties of their procedure are satisfactory with sample sizes of 500 (or more) observations.

The paper by Hidalgo and Kreiss discusses goodness-of-fit tests for linear covariance-stationary processes which include long-memory processes with $d<\frac{1}{2}$. The proposed tests are set in the frequency domain and involve the spectral distribution function, the integral of the spectral density function. However, the limiting distribution of the proposed statistic is not pivotal and depends on unknown nuisance parameters. The authors show the validity of a frequency domain bootstrap which consists of resampling the periodogram. This procedure is natural given the test statistic used and is valid since the periodogram computed at different frequencies is independent. A MC experiment shows that test size is well controlled in finite samples. However, power is quite low for persistent alternatives.

In their paper, Horowitz, Lobato, Nankervis, and Savin consider the problem of testing the null hypothesis that the first K autocorrelations of a covariance stationary time series are zero, even if other forms of dependence (such as nonlinear dependence) may be present. The test statistic is the Box–Pierce based on the first K autocorrelations. Since the null hypothesis allows for statistical dependence, the test statistic is not asymptotically pivotal. To deal with this problem they propose using a double block-of-blocks bootstrap with prewhitening, which appears to be superior to the block or stationary bootstrap. The authors also present a simulation experiment, which shows that the performance of the procedure is quite satisfactory.

Finally, the paper by Li and Tcakz presents a new test for evaluating conditional density functions for time-series data. The test statistic is based on the integral of the weighted squared distance between a kernel estimate of the conditional density function and the conditional density under the null hypothesis. The authors show that this statistic has a limiting standard normal distribution under the null hypothesis and diverges under the alternative hypothesis, yielding a consistent test. As with many such specification tests, the asymptotic distribution does not provide reliable inference in samples of reasonable size. Moreover, the results are sensitive to the choice of smoothing parameter in the non-parametric kernel estimation of the conditional density. Thus, the authors rely on the bootstrap to improve the behavior of the test. Simulation results show good size control for all choices of the smoothing parameter considered. The model used is an Ornstein–Uhlenbeck process as suggested to the interest rate model of Vasicek (1977). Power is analyzed by generating data from a square-root process suggested as an alternative to the Vasicek model of Cox et al. (1985). Power is good for samples as small as 50 and is not very sensitive to the choice of smoothing parameter. The authors use the new test to reject two models of inflation and conclude that monetary policy may be misguided if based on incorrect parametric assumptions.

Before closing, we want to take this opportunity to thank the editors of this journal, in particular Cheng Hsiao, for their support and advice. Financial support for the conference and the preparation of this special issue was provided by the Centre interuniversitaire de recherche en économomie quantitative (CIREQ), the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Mathematics of Information Technology and Complex Systems (MITACS) Network of Centres of Excellence, the Social Science and Humanities Research Council of Canada, the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Institut de finance mathématique de Montréal (IFM2), the Centre interuniversitaire de recherche en analyse des organisations (CIRANO), the Natural Sciences and Engineering Research Council of Canada, the Fonds de recherche sur la société et la culture (Québec), and the Fonds de recherche sur la nature et les technologies (Québec). Specials thanks go to the referees who devoted time and energy to the improvement of the papers in this volume. Finally, the organizational skills of the staff at CRDE and CIREQ were invaluable in the organization of the conference and the preparation of this issue. We want to express our gratitude to Sharon Brewer Chantal Brunet, Marie-Christine Thirion, and Francine Martel.

References

Aït-Sahalia, Y., Duarte, J., 2003. Nonparametric option pricing under shape restrictions. Journal of Econometrics 116, 9–47.

Andrews, D.W.K., 1997. A conditional Kolmogorov test. Econometrica 65, 1097–1128.

Barnard, G.A., 1963. Comment on 'The spectral analysis of point processes' by M.S. Bartlett. Journal of the Royal Statistical Society, Series B 25, 294.

Basawa, I.V., Mallik, A.K., McCormick, W.P., Reeves, J.H., Taylor, R.L., 1991. Bootstrapping unstable first-order autoregressive processes. The Annals of Statistics 19, 1098–1101.

Bertail, P., Politis, D., Romano, J., 1999. On subsampling estimation with unknown rate of convergence. Journal of the American Statistical Association 94 (446), 487–508.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–654.

Cowles, A., 1933. Can stock market forecasters forecast? Econometrica 1 (3), 309–324.

Cox, J., Ingersoll, J., Ross, S., 1985. A theory of the term structure of interest rates. Econometrica 53, 385–408.

Davidson, R., MacKinnon, J.G., 1981. Several tests for model specification in the presence of alternative hypotheses. Econometrica 49, 781–793.

Davidson, R., MacKinnon, J.G., 1999. Bootstrap testing in nonlinear models. International Economic Review 40, 487–508.

Davidson, R., MacKinnon, J.G., 2000. Improving the reliability of bootstrap tests. Technical Report, Department of Economics, Queen's University Kingston, Ontario.

Diebold, F.X., Gunther, T., Tay, A.S., 1998. Evaluating density forecasts with applications to finance and management. International Economic Review 39, 863–883.

Dwass, M., 1957. Modified randomization tests for nonparametric hypotheses. Annals of Mathematical Statistics 28, 181–187.

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

Hall, P., 1992. The Bootstrap and Edgeworth Expansion. Springer, New York.

Hall, P., Horowitz, J.L., 1996. Bootstrap critical values for tests based on generalized-method-of-moments estimators. Econometrica 64, 891–916.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y., 1992. Testing for the null hypothesis of stationarity against the alternative of a unit root. Journal of Econometrics 54, 159–178.

Lettau, M., Ludvigson, S., 2001. Consumption aggregate wealth and expected stock returns. Journal of Finance LVI(3), 815–849

Li, H., Maddala, G.S., 1995. Bootstrapping cointegrating regressions. Journal of Econometrics 80, 231–254.

Park, J.Y., Phillips, P.C.B., 1988. Statistical inference in regressions with integrated processes: Part 1. Econometric Theory 4, 468–497.

Politis, D.N., Romano, J.P., 1994. The stationary bootstrap. Journal of the American Statistical Association 89, 1303–1313.

Saikkonen, P., 1991. Asymptotically efficient estimation of cointegration regressions. Econometric Theory 7, 1–21.

Schwert, G.W., 1989. Tests for unit roots: a Monte Carlo investigation. Journal of Business and Economic Statistics 7, 147–159.

Sowell, F., 1992. Maximum likelihood estimation of fractionally integrated time series models. Journal of Econometrics 53, 165–188.

Stock, J.H., Watson, M.W., 1993. A simple estimator of cointegrating vectors in higher order integrated systems. Econometrica 61, 783–820.

Swensen, A.R., 2003. Bootstrapping unit root tests for integrated processes. Journal of Time Series Analysis 24, 99–126.

Vasicek, O., 1977. An equilibrium characterization of the term structure. Journal of Financial Economics 5, 177–188.

Jean-Marie Dufour[a,b,c], Benoit Perron[a,b,c]
[a]*CIRANO, Canada*
[b]*CIREQ, Canada*
[c]*Département de sciences économiques, Université de Montréal, Canada*
E-mail address: jean.marie.dufour@umontreal.ca (J.-M. Dufour)
URL: http://www.fas.umontreal.ca/SCECO/Dufour