



Time Series Inference Applications of R in Finance and Econometrics

R User Conference, useR! 2010, National Inst. of Standards
& Technology, Gaithersburg, MD, July 21-23, 2010

H. D. Vinod

Professor of Economics,
Fordham University, New York **For a copy of the paper write to:**
vinod@fordham.edu

OUTLINE

- 1 Time Series economists failed to predict 2008 Recession. Leamer (JEP, 2010“Tantalus on the road to Asymptopia”) blames excess math. Wiener-Kolmogorov-Khintchine (WKK) 1930's inference tools unrealistic, assume $I(0)$.
- 2 ‘meboot’ bootstrap Conf Intervals admit $I(d)$ for any d , evolving irreversible dynamics, simultaneity, structural changes, nonlinearities and idiosyncratic markets.
- 3 Spurious regression (meboot vs block bootstrap). Use ‘ConvergenceConcepts’ package and meboot avoid arbitrary/ weak instruments (GMM) for finite n .
- 4 meboot two-way map time domain \leftrightarrow values domain suggests Hetero. correction GLS not ‘WhiteWash’ of HAC standard errors. Flexible inference with Kernel & meboot.
- 5 Better forecasts use Exploratory Data Analysis(EDA) in R.

'meboot' Maximum Entropy Bootstrap Motivation

Efron's bootstrap known to fail for non-iid data. Vinod (2004, 6,8) offers a **free 'meboot' package in R**. (1) Statistical time series inference 1930's has an infinite population of time series (=ensemble Ω). $I(0)$ model unsuitable for short evolving time series. (2) Many disadvantages of converting data into stationary data before modeling starts. Differencing changes the specification! (3) Changing definitions are generally not a problem in natural sciences. But, quality and content of GDP, names of stocks in the Dow Jones (DJIA) etc. change. Since $n > 100$ series problematic, asymptopia is unattainable. Basic 'meboot' **trick: joint sorting** two columns of a matrix to achieve: t-domain \leftrightarrow values domain maps. Proof of the pudding... We illustrate with airline passenger data to create Ω ensemble of J time series for inference.

Steps in Vinod's ME bootstrap to create a replicate of x_t .

- 1 Sort the original data in increasing order. Storing the ordering index vector allows t-dom \leftrightarrow v-dom maps.
- 2 Compute intermediate points on the sorted series.
- 3 Compute the mean of ME density within each interval & *mean preserving constraint* (ergodicity) satisfied.
- 4 Generate random numbers from the $[0,1]$ uniform interval and compute sample quantiles at those points.
- 5 Apply to the sample quantiles the correct order to keep the dependence relationships of the observed data using ordering index of step 1.
- 6 Repeat steps 2 to 5 several times (e.g. $J=999$).

Example: AirPassengers evolving monthly time series of Total airline passengers, 1949 to 1960. Figure depicts data and 49 'meboot' replicates. **Usual iid boot samples from the Empirical CDF**. ECDF resides in the values-domain **iid boot fails for time series** due to absence of map (v-dom) \rightarrow (t-dom) and No Shape Retention! **'meboot' retains both ACF and periodogram shape** properties of data series in (t-dom) without imposing parametric constraints. See shape retention in next figure.

'meboot' and two-way maps have many applications in Econometric estimation and inference using its J (=999 say) **replicates of any evolving series**. Both iid and 'meboot' allocate discrete mass ($1/T$) near each observation.

Problems with iid boot, all solved by ‘meboot’

- 1 iid boot excludes nearby values. For example, let $x_t = 49.2$. Since 49.19 or 49.24, both of which round to $x_t = 49.2$, there is no justification for excluding all such values.
- 2 resamples must lie in the closed interval $[\min(x_t), \max(x_t)]$.
- 3 any dependence information in sequence $(x_1, \dots, x_t, x_{t+1}, \dots, x_T)$ is lost in the shuffle.

Vinod and Lopez-de-Lacalle explain ‘meboot’ with examples (including panel data). Mass and mean preserving constraints on ME density. <http://www.jstatsoft.org/v29/i05>
Satisfy ergodic theorem (time avg=ensemble avg).

'meboot' uses $J = 999$ estimates of $(b_i^* - b_i)$ to approximate the sampling distribution of $(b_i - \beta_i)$. Here in the land of finite n , I suggest using 'meboot' and the 'ConvergenceConcepts' package to assess relevant convergence properties.

Example: testing Friedman's permanent income hypothesis.

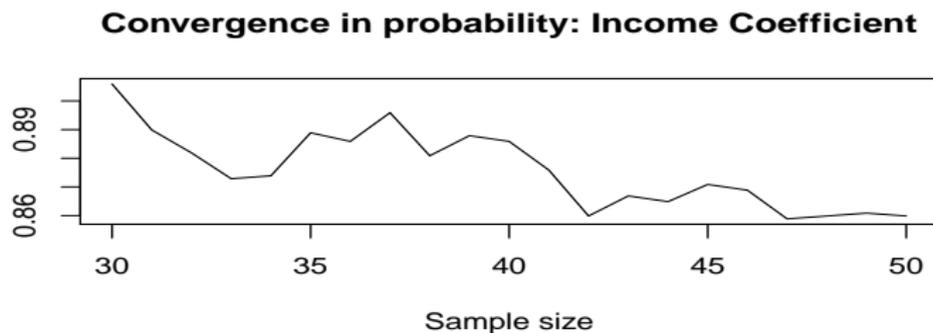
$$C_t = \beta_1 + \beta_2 C_{t-1} + \beta_3 Y_{t-1}$$

If PIH then β_3 should be insignificant. 'meboot' has 'checkConv' function to simulate with known β_i . Figures show that $\beta_3 \rightarrow 0$ near last 10% of n .

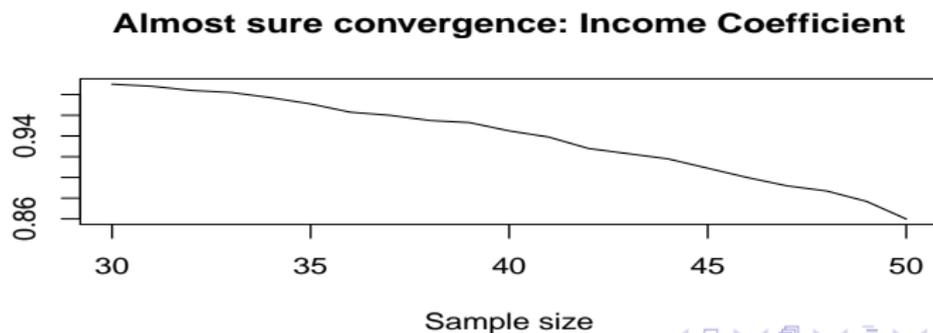
More generally, what is the effect of consumer confidence, inflation, wealth-effect from real estate and stock markets, etc. on C_t ? 'meboot' simulations can help understand the effect of confounding (additive and interactive) variables based on a variety of control variables.

Figure: Checking 'convergence in probability' and 'almost sure' convergence for a simulated consumption function

Criterion using 999 sample paths



Criterion using 999 sample paths



Kernel Regr

Consider specification of models with (confounding) control variables. Since functional forms are uncertain, 'meboot' allows inference in these hard problems when n is small.

Kernel regression estimates flexible function (NO assumptions):

$$y = R(x) + \epsilon,$$

$$R(x) = \frac{\sum_{t=1}^T y_t K(w_t)}{\sum_{t=1}^T K(w_t)}, \text{ where } K(w_t) \text{ kernel, weights } w_t = \frac{(x_t - x)}{h},$$

evaluated at x .

The amorphous partial derivative (apd) of $R(x)$ w.r.t. j th

$$\text{regressor } \text{apd}(x_j) = \sum_{t=1}^T y_t (K_{1t} - K_{2t}),$$

$$K_{1t} = \frac{K'}{S_K}, \quad K_{2t} = \frac{K S_{K'}}{(S_K)^2}, \quad K' \text{ is derivative } S_K \text{ is } \sum_{t=1}^T K, \text{ and}$$

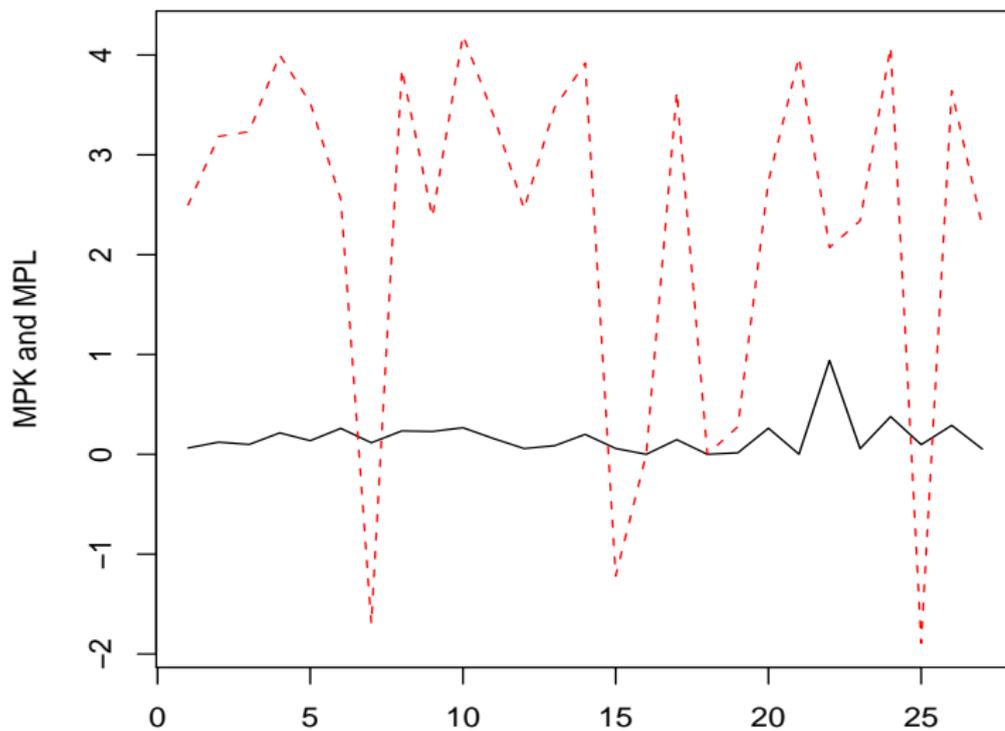
$$S_{K'} = \sum_{t=1}^T K'.$$

The R package “np” has ‘local linear estimator,’ fitted values \hat{R} and the gradient, $\partial\hat{R}/\partial x$, evaluated at all T evaluation points x . Vinod (2008) text suggests the following modification:

$$\text{apd}'(x_j) = \text{mean}(\partial\hat{R}/\partial x_j), \quad SE(\text{apd}') = \text{sd}(\partial\hat{R}/\partial x_j), \quad (1)$$

where SE=std error, sd=std dev. Since the apd' of (1) is a simple average of T gradients, it benefits from the central limit theory. Some x_j can be control variables. ‘meboot’ permits confidence intervals **CI95** around estimated apd's. Then one can determine if control variables are **confounding** and if deeper relations are significant. See code in my "Hands-on Intermediate Econometrics Using R" pub. by World Scientific.

Marginal productivity of capital, (solid line) and labor



EDA

Failure to predict great recession of 2008 shows a need for fresh thinking and search for new relations.

Economists rely too much on theory and too little on data.

Why not use Exploratory Data Analysis (EDA) in R? For example, package 'UsingR' has 'simple.eda' to plot histogram, boxplot and normal quantile-quantile plot. Use R graphics tools help assess potential relations to help formulate new hypotheses for more reliable forecasts/ theories.

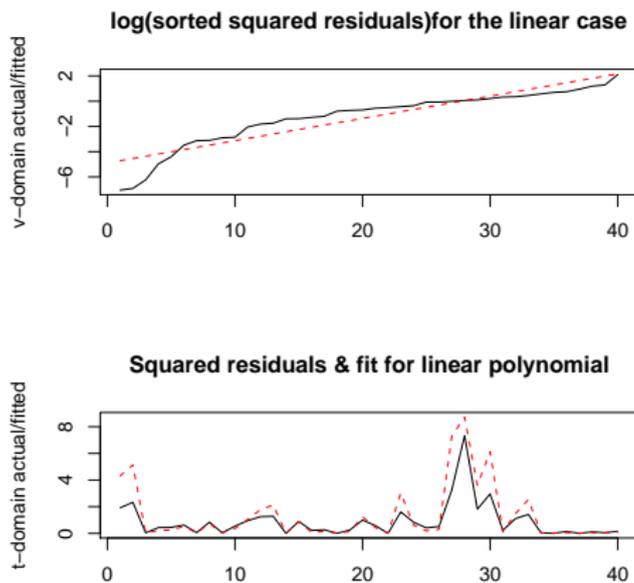
New R tools for multivariate EDA include methods for GLMM (Generalised Linear Mixed Model), MARS (multivariate additive regression splines).

GLS

- 1 My paper in "Advances in Social Science Research Using R," Springer, New York, <http://www.springer.com/statistics/business%2C+economics+%26+finance/book/978-1-4419-1763-8> explains efficient est. avoiding White's HAC standard errors.

GLS correction for heteroscedasticity uses sorted squared residuals (v-dom) and regress them on **sequence of numbers $t=(1,2,\dots)$** . The upper panel shows how only two parameters in $(a + bt)$ can represent hetero. in (v-dom). Reverse map in (t-dom) in lower panel shows that it is effective and allows correction by GLS. A simulation shows improved efficiency.

Figure: Heteroscedasticity modeling using $a + bt$, v-dom upper panel, t-dom lower panel shows good fit with two parameters a, b .



CONCLUSION (Beyond WKK Asymptopia)

- 1 'meboot' t-dom \leftrightarrow v-dom maps help retain SHAPE, satisfies ergodic theorem. Allow small sample inference with problem-specific convergence checks without transforming short irreversible series into $I(0)$ stationary series. Solves spurious regression in finite samples. Potentially useful in all kinds of simulations including 'calibration' etc. New tools for GLS solve the "White-wash" problem.
- 2 Flexible functional forms in relations with Kernel estimates of partials (apd's) analyze quantitative contribution of a regressor (control or focus) on the dependent variable with 'meboot' 95% CI.
- 3 Many powerful R tools and graphics including EDA should help prevent future forecasting failures.

Bibliography

-  Vinod, 2004. Ranking Mutual Funds Using Unconventional Utility Theory and Stochastic Dominance. *J of Emp. Finance* 11(3), 353-377.
-  Vinod, 2006. Maximum entropy ensembles for time series inference in economics. *J of Asian Economics*, 17(6), 955-978.
-  Vinod, & Lopez-de-Lacalle, 2009. The ‘meboot’ R package. *J of Stat Software*, 29(5)
<http://www.jstatsoft.org/v29/i05>
-  Vinod 2010. Superior estimation and inference avoiding heteroscedasticity & flawed pivots, in *Advances in Social Science Research Using R*. Springer, NY 39–63.
Volcker Rule <http://ssrn.com/abstract=1597450>
Spurious Regression [abstract=1560074](http://ssrn.com/abstract=1560074), New
Keynesian Phillips Curve [abstract=1560622](http://ssrn.com/abstract=1560622)