<u>REIN TAAGEPERA</u> – 11/19/09 COLLOQUIUM

Department of Political Science, UCI

Title: <u>Interconnected knowledge requires symmetric regression</u>

One thing that impresses me about physics is networks of interlocking quantitative relationships. They depend on algebraic equations, which are reversible and transitive, so that that any variable or parameter in an equation can be calculated when the rest are given. For equations connecting x, y and z pairwise, the same value of z is obtained whether one calculates z from x or first calculates y from x and then z from y.

In contrast, linear OLS regression equations widely used in social sciences are directional and non-transitive. Regression lines y-on-x and x-on-y differ, so that one should properly use y <-- $a+bx$ rather than $y=a+bx$. Regressing y on x and then z on y yields values different from z regressed on x directly. Hence OLS regression cannot lead to interlocking relationships. In particular, suppose a predictive model holds, but with appreciable scatter. When predicted values are regresssed against the actual, the slope comes out as $dP/dA<1$; but reverse regression yields a contradictory $dA/dP<1$.

Scale-independent symmetric linear regression exists, reversible and transitive, and it should be used when the goal is establishing general relationships rather than some applied purpose. R-squared reflects lack of scatter around this symmetric regression line, while the two OLS slopes are mixed measures of symmetric slope (B) and scatter: $b=B|R|$ and $b''=B/|R|$, respectively. The larger the scatter, the more the OLS lines diverge from true slope