

## **Surprising insights from regression analysis when explanatory variables are percentages**

Ulrich B. Morawetz and H. Allen Klaiber

### **Abstract:**

Regression analysis in economics, sociology and political science often use explanatory variables measured in percentages (e.g., % agricultural land and % forest of a farm). To avoid perfect multicollinearity with the constant of the model one of the components needs to be dropped (e.g., % forest). The estimated coefficient is then interpreted as the effect of a 1%-point increase of the component that remained in the model (e.g., agricultural land) at the cost of the dropped component (e.g., a 1%-point reduction in forest). We show that this interpretation is in general only correct if the number of components is limited to two. Otherwise, the estimated coefficients can be influenced by the correlation structure between components and their partial effects, and interpretation is possible only for specific data generating processes. Furthermore, we show the conditions under which a Two Stage Least Square regression can be applied if one of the components is endogenous.