demand side, can only be analyzed by means of case studies and by an investigation of the institutional arrangements. Chapters 4 and 5 will provide more evidence on these important supplements to the multiple regression analysis.

Data scarcity and reliability are, of course, a problem for studies like this one. It has already been mentioned that there may be a slight bias due to the unequal distribution of missing data with regard to country size. Clearly, data for smaller countries are more difficult to obtain, and many VSC do not even provide national accounts. Nevertheless, our sample includes data for twelve countries which have fewer than a million inhabitants (see Table A.5). VSC are hence underrepresented in comparison to larger countries, but they are sufficiently represented to allow us to draw conclusions from the results.

Data scarcity (especially for VSC) and reliability are related problems, and there are two approaches to coping with them. First, the approach followed in this study: One can be aware of the shortcomings and the limitations of the results, but, nonetheless, stick to the empirical analysis to get a broad-stroke picture of the structure of the question at hand. Second, one can stress the caveats and abstain from a quantitative assessment, which, as a consequence, means that it is not possible to answer the question concerning the relationship between country size and government size generally.

As is well known in econometrics, when cross-section data of countries are in use, there is often a problem of heteroscedasticity involved.[53] Heteroscedasticity is a violation of one of the central assumptions of the classical linear regression model, which states that the variance of each disturbance term ui should be constant. Symbolically, $E(u_i^2) = \sigma_i^2$ (heteroscedasticity) instead of the assumption $E(u_i^2) = \sigma^2$ (homoscedasticity). The consequence of heteroscedasticity is that the estimated coefficients no longer comply with the criterion of minimum variance in the class of linear unbiased estimators; hence they are no longer BLUE (best linear unbiased estimator). It is hazardous or misleading to make inferences based on OLS (ordinary least squares) estimations in the presence of heteroscedasticitiy.

---

[53]   Details on heteroscedasticity, detection methods and remedial measures can be found in almost any standard textbook on econometrics.