

Model Building: The Three “Tiers” of Variables

Thanks to convenient, high speed statistical computing software, it is easier than ever to experiment with empirical models by adding more and more variables in various combinations. Too often, inexperienced researchers keep trying “stuff” just to see what happens. Not only does this lead to random findings of pseudo results, it can create complex intercorrelations and causal complexities that make model interpretation hazardous.

To simplify the process of model building, it is useful to classify variables into three tiers of importance:

1) Tier 1: Main variables

These are the variables that the client is most interested in. There are usually just 2 or 3 main variables. These should be included in every model. The focus of your research is the robustness of the results you obtain for these variables

2) Tier 2: Control variables and permutations/combinations of the main variables

These include the most important control variables. You may also try exponents, interactions or other permutations of the main variables. You may obtain some nuanced results concerning the main variables as well as look for robustness. It is okay to experiment a bit here, but not too much.

3) Tier 3: Permutations/Combinations of the control variables

Using exponents, interactions, and other permutations of control variables can get you into trouble. There is no great interest in the findings with tier 3 variables and often you are trying to obtain nuances of results that were not very strong to begin with (in tier 2). Best to keep these to a bare minimum, if you use them at all, and use them mainly to test for robustness of the main results.