Description for data used in Phiromswad and Hoover, “Selecting Instrumental Variables: A Graph-theoretic Approach”
20 August 2013

The set of potential growth determinants is constructed as a non-overlapping set of variables studied in Doppelhofer, Miller, and Sala-i-Martin (2004) and Levine and Renelt (1992). In addition the secondary enrollment rate in 1960, which was inexplicably omitted from the data set of Doppelhofer et al. is added from Sala-i-Martin’s (1997) data set. In total, our data set includes 87 variables. Countries for which observations on any of the variables were missing are not included in our data set, leaving 72 countries.

The file growth_IVD.xls contains the actual data used in the paper. The worksheet “info” contains short description of all variables as well as the variables’ code from the original source.

In addition, the worksheet “info” also documents the background knowledge information we used in the Stage 1 of the IVD algorithm. In particular, the background knowledge can be classified into four types.

The first type is geography, natural resource and regional dummy variables. I assume that variables of this type are predetermined relative to all other variables (including variables within this group). At the beginning of step 1 of the IVD algorithm (the elimination step), undirected edges between each variable in this group are removed. At the beginning of step 3 of the IVD algorithm (the orientation step),
all undirected edges with variables in this group are oriented away from variables in this group.

The second type is colonial origins variables. I assume that variables in this group are predetermined relative to all variables (including variables within this group) except for geography, natural resource, and regional dummy variables. At the beginning of step 1 of the IVD algorithm (the elimination step), undirected edges between each variable in this group are removed. At the beginning of step 3 of the IVD algorithm (the orientation step), all undirected edges with variables in this group are oriented away from variables in this group (except undirected edges with geography, natural resource and regional dummy variables).

The third type is religious variables (which are measured at 1900). I assume that variables in this group are predetermined relative to all other variables except the two types discussed earlier. At the beginning of step 3 of the IVD algorithm (the orientation step), all undirected edges with variables in this group are oriented away from variables in this group (except undirected edges with the two types discussed earlier).

The fourth type is variables measured at the beginning of 1960s. I assume that variables in this group are predetermined relative to all other variables except the three types discussed earlier. At the beginning of step 3 of the IVD algorithm (the orientation step), all undirected edges with variables in this group are oriented away from variables in this group (except undirected edges with the three types discussed earlier).