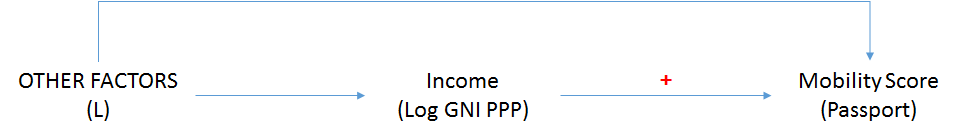
Mahesh note on interpreting GNI

A quick response: your interpretation of the graph is exactly right. There is a positive correlation between gross national income, an indicator of economic performance, and mobility, as measured by the passport mobility scores. As you pointed out, there is also significant variation across income levels that do not explain variation in passport mobility. It looks like the variation is also increasing with increasing log GNI, i.e. the greater the log GNI, the greater the passport mobility on average (the red line), but this is also accompanied by greater variation in passport mobility, which can be seen by the fact that the dispersion in the gray dots away from the red line (in stats, the variance of the residuals) gets larger as GNI increases. For countries with large GNI, for example, some countries have very high mobility, but others have very low mobility, relative to countries with low GNI. In other words, the variance of the error looks to be heteroskedastic - non-constant variance.

Speaking about the variance is one thing. But an equally important message from a structural modeling perspective is that this is a correlation and not a causal story. The reason why you might be seeing the positive correlation is because there may be many other omitted factors (L in the graph below) that are driving both income as well as mobility scores, and it may be these factors that are explaining the positive correlation. For example, countries with better institutions (institutions are the omitted unmeasured factor L) are likely to have higher incomes (so the arrow between institutions L and income is positive), and countries with better institutions are also more likely to grant their citizens with higher mobility independently of income (so the arrow between L and mobility is positive). If this were true, then we would be likely to overestimate the correlation between income and mobility because (a possibly significant) part of that correlation may be driven by institutional quality, which is unmeasured.



Regarding interpretation: the easiest way to interpret logged variables is in terms of percent changes as opposed to absolute changes.

So, for example: if you are running the regression:

Mobility\_c = alpha + beta\*GNI\_c + error\_c

where Mobility\_c is mobility for a country c (measured by the score), alpha is the constant, beta is the slope, GNI\_c is GNI in **absolute levels of income (measured in $)**, and error\_c is the error term, then the interpretation of the slope would be that a $1 increase in GNI is associated with a beta unit increase in the mobility score, keeping everything else constant.

If, however, you run the regression:

Mobility\_c = alpha + beta\*log(GNI\_c) + error\_c

where log(GNI\_c) is the **log level of income (measured in log $)**, then the interpretation of the slope would be that a **1 percent** increase in GNI is associated with a beta unit increase in the mobility score, keeping everything else constant.