How to use proxy indicators in predictive analytics

In forecasting the use of proxies often happens unconsciously

When the weather is miserable, I discover myself longing for next days' weather forecast. It can't be worse, but maybe it will be better. Reflecting upon this I realized the relevance of your starting point when you do a forecast. Your starting point is often a valuable predictor. Today it has been raining cats and dogs. In the absence of any other data, for predicting tomorrow's weather the safest forecast is to predict more cats and dogs, right? When doing so we take today's weather as a proxy indicator for forecasting tomorrow's weather. Here the proxy indicator has a causal link with the prediction. The fact that today has been rainy may relate to a depression that tomorrow will still be there (or rather: here) and that may bring more rain.

During my strategic forecasting work, I indeed usually observe a direct causal relation between the proxy indicator and the to be predicted phenomenon. The following examples come to mind:

- population size predictions as proxy indicator for future food consumption. In global dairy consumption, half of the global volume growth is simply and over the years consistently attributable to population growth.

- oil price predictions as proxy for future consumer spending power in resource-rich emerging market economies.

There is a common pattern in these examples. We choose a prediction we know we can get hold of or that we can produce ourselves with a decent reliability – or so we wish to believe. We subsequently use it to (quantitatively) predict a phenomenon we do not yet have a forecast for but the prediction of which we are interested in.

Attributes of good proxy indicators

Predicting future population sizes of e.g. a country is usually among the most reliable predictions one can think of. This is true because countries are usually so large that an individual birth does no longer matter to the total number of newborns in a given period. To paraphrase Stalin: the birth of an individual is worth a celebration, the birth of a million is a statistic. What prevents inaccuracies is that e.g. population growth has only a modest dependency on other factors like e.g. economic conditions.

With a prediction in hand, the only thing we need is a clear historic relation between the phenomenon to be predicted and our proxy indicator – e.g. the population size. Having the population size forecast enables us to predict the phenomenon we look for through forward extrapolation of the historic relation with the population size prediction as proxy indicator.

By now we have implicitly obtained three attributes for a good proxy indicator. Such indicator is:

- based on phenomena governed by the law of the large numbers like population growth[1].

- widely available and possibly estimated by enough different reputable parties to enable to preferably come to a 'consensus estimate' based on different sources.

- having a strong historic correlation with the indicator that we want to predict, which we have no reason to doubt will also apply in the future.

A final warning

For all their usefulness the value of proxy indicators should not be overrated. The example below I guess powerfully conveys that message. Last year, an Indian friend treated me to a magnificent dinner in Hyderabad (India) on the terrace of what formerly was a splendid local prince's palace. When the first dish was being served, he ignored all silver cutlery and started to eat the food with his hands. Having lived in Asia I knew that habit, but I found it rather peculiar given the plush surroundings we were in. When he noticed my unspoken surprise, what he subsequently said has stayed with me ever since:

"Eating with cutlery is like making love through a proxy."

Notes:

.1.   Taleb, N.N. [2008], The Black Swan – the impact of the highly improbable, Penguin Books, London.