Description of the Beyer and Wacker (2022) data set

Overview

This note briefly explains the dataset by Beyer and Wacker (2022) that updates the dataset by Araujo et al. (2016). Furthermore, it provides a very quick overview about the econometrics of growth regressions (section 2) and a reference command how they are implemented in STATA with the mentioned data set (section 3).

When using the data, please cite it as:

> Beyer, Robert C.M. and Konstantin M. Wacker (2022): "Good Enough for Outstanding Growth: The Experience of Bangladesh in Comparative Perspective." World Bank Policy Research Paper 10150.

1. Variables and data sources

The data set contains data on income levels and potential covariables for ~150 countries over the time period 1970-2019, averaged over non-overlapping 5-year periods *t*, i.e. the first period covers the average over the years 1970-1974, while the last period (*T=10*) covers the average over the years 2015-2019.

The key dependent variable one aims to explain is *lrgdpna_pc*, the real (inflation-corrected) income level per capita, in logarithmic terms. The variable is taken from PWT10.0 and constructed by dividing *rgdpna* by *pop*. The natural log is taken such that changes in the explanatory variables can be interpreted as percent changes in p.c. income levels. Although national income data are converted into (constant 2017) US$, note that this variable does not allow for proper comparison of welfare across countries because it does not account for differences in price levels across countries. Such comparisons are facilitated through PPP adjustments, as in the PWT series *rgdpo* or *rgdpe* (see Feenstra et al., 2015), which are available upon request.

The table on the next page gives an overview about the variables contained in the data set.

Name	Description	Source
lrgdpna_pc	Log of real GDP per capita (at constant 2017 US$)	PWT10.0
lhc	Log human capital index (education)	PWT10.0
lkg	Log government consumption	PWT10.0
lrer	Log real exchange rate (pl_gdpo/xr)	PWT10.0
ltraderesid	Log of trade [(exports+imports)/GDP] that is unexplained by exporter population size and global demand	Own estimation based on PWT10.0
lcredit	log of domestic credit to to the private sector as a percentage of GDP	WDI
linflation_na	Log of (inflation+1)	PWT10.0
lphoneline	Log fixed telephone line	WDI
infrastructure_index	Self-constructed index which is a weighted average of phone lines, mobile phones, internet connections, electricity access, and secure internet connections.	Own calculation based on WDI
d.ltot	Changes of log terms of trade d.[ln(pl_x/pl_m)]	PWT10.0
lEDI_ipol	Log of export diversification index, mostly based on Papageorgiou et al. (2015).	IMF and UNCTAD
lFDIstock_ipol	Log of FDI stock relative to GDP	UNCTAD
actotal	Variable indicating the magnitude of political violence	PolityV by Systemicpeace.org
dum_mepv	Dummy variable indicating that time period covers a mayor episode of political violence.	PolityV by Systemicpeace.org
fincrisis_years	Fraction of years in financial crisis (banking, currency, or sovereign debt).	Laeven and Valencia (2020)
dum_fincrisis	Dummy variable indicating if some financial crisis has occurred during period	Laeven and Valencia (2020)
sd_temperature	Standard deviation of monthly mean temperatures (in Celsius)	CCKP
sd_growth	standard deviation of growth over a 5-year period (lagged by 2 years).	PWT10.0
urbanpop	Share of urban population	WDI
emprate	Employment rate, i.e., the fraction of population that is in employment (emp/pop).	PWT10.0
remittances	Remittances received, as percent of GDP	WDI
gini_mkt	Gini coefficient of market incomes	Standardized World Income Inequality Database (Solt, 2019)
period	Identifier for 5-year periods. Period 1 indicates average of years 1970-1974, period 2 is 1975-1979 etc.	
geo	Panel identifier for country	

2. Short introduction to growth econometrics: the neoclassical growth model

The standard approach in the empirical growth literature is to model the logarithm of a country c's per capita income level at period t, ln y_{ct}, as a function of growth correlates, $x_{k,ct}$ ($k=1,\ldots,K$), and the lagged income level, ln $y_{c,t-1}$ (e.g. Araujo et al., 2016; Moller and Wacker, 2017; Brueckner and Lederman, 2018):

$$\ln y_{ct} = \theta \ln y_{c,t-1} + \beta_1 x_{1,ct} + \cdots + \beta_K x_{K,ct} + \delta_t + \alpha_c + \varepsilon_{ct}. \quad (1)$$

This equation includes an autoregressive term ln $y_{c,t-1}$ ('lagged dependent variable') that determines the dynamics of the model, country fixed effects α_c, which account for (time-invariant) unobserved differences in income levels across countries, and period fixed effects δ_t, which account for global shocks to income levels (e.g. during the global financial crisis).

Taking first differences of equation (1) with respect to time gives:

$$\underbrace{\Delta \ln y_{ct}}_{acutal\ growth} = \underbrace{\theta \Delta \ln y_{c,t-1}}_{persistence} + \underbrace{\beta_1 \Delta x_{1,ct} + \cdots + \beta_k \Delta x_{k,ct}}_{innovations\ in\ growth\ correlates} + \underbrace{\Delta\delta_t + \Delta\varepsilon_{ct}}_{residual}, \quad (2)$$

where Δ is the difference ('lag') operator: $\Delta x=x_t-x_{t-1}$. Note that log-changes approximate percent changes such that equation (2) is a 'growth equation'. Country fixed effects are 'differenced away' after taking first differences since they are time-invariant but the other central model parameters $\theta, \beta_1, \ldots, \beta_K$ are mathematically equivalent in equations (1) and (2). Because growth can be rather volatile over time (e.g. Easterly et al., 1993; Pritchett and Summers, 2015), estimation of the model in levels usually provides more reliable estimates of those parameters, which can then be applied to the analysis of growth in the first-difference equation (2).

Equations (1) and (2) imply that an innovation in a growth correlate x_k has a permanent effect on the income level but a temporary effect on the growth rate. A change in x_k, Δx_k, in period t will have a contemporaneous effect β_k on the growth rate. Since this growth rate, $\Delta \ln y_t$, will be the lagged growth rate in the next period $t+1$, the effect shows some 'autoregressive' persistence θ over time. However, if $\theta<1$, which is usually the case, the effect fades out over time. This is economically intuitive: if we build an infrastructure project today, it will boost growth and the income level. Some positive growth effect may still be present today but it will fade out over time. Equation (2) implies that the growth impulse in period $t+j$ will be equal to $\theta^j\beta_k$. The sum of this geometric series converges to $\beta_K/(1-\theta)$, such that the short-run effect of a change in x_k on income levels, $\partial \ln y_t/\partial x_{k,t}$ equals β_k, while the long-run effect is given as:

$$\frac{\partial \ln y_{t+j}}{\partial x_{k,t}} = \beta_k + \theta\beta_k + \theta^2\beta_k + \theta^3\beta_k + \ldots + \theta^j\beta_k \underset{j\to\infty}{\Longrightarrow} \frac{\beta_k}{(1-\theta)}.$$

In the presence of country fixed effects, the lagged dependent variable governs a country's convergence towards its own steady-state. This can be seen after subtracting ln $y_{c,t-1}$ from both sides of equation (1), which yields:

$$\Delta \ln y_{ct} = (\theta - 1)\ln y_{c,t-1} + \beta_1 x_{1,ct} + \cdots + \beta_K x_{K,ct} + \delta_t + \alpha_c + \varepsilon_{ct}. \quad (3)$$

Suppose that a stochastic income shock $\varepsilon_{c,t-1} > 0$ has happened in period *t-1*, which was not driven by any fundamental growth correlates *x*. This will increase ln $y_{c,t-1}$ but since $(\theta\text{-}1)<0$, it will lower the growth rate $\Delta \ln y_{ct}$ in the subsequent period, such that the economy converges back towards its steady state, as the neoclassical Solow model implies. In a cross-country setup, where no country fixed effects are present, the coefficient θ governs beta convergence across countries. In such a cross-country context, $(\theta\text{-}1)<0$ implies that an economy with a lower initial income level grows faster than an economy with a higher initial income level (see Patel et al., 2021). Finally note that the dynamic implications of this model are different from those in endogenous growth models, where, for example, education and research or institutional changes can have permanent effect on the growth rate (by speeding up technological progress).

3. Estimation in STATA and reference results

Several approaches are possible to estimate equation (1). One straight-forward option is to use a simple 'fixed effect' approach, which is an ordinary least squares (OLS) estimation augmented with dummy variables for panel units (and possibly time periods), hence also called 'least square dummy variable' (LSDV) estimation. In STATA, one way to implement such an estimation is:

```
xtreg lrgdpna_pc L.lrgdpna_pc lkg lrer ltraderesid lcredit linflation infra_index d.ltot lEDI_ipol lEDI_ipol_sq lFDIstock_ip actotal dum_finc lhc d.gini i.period, fe robust
```

Note that "L." is the 'lag operator' (taking the first lag of a variable) and that "d." is the 'first difference operator' (taking first differences with respect to time). "i." indicates that the following variable (period) should be transformed into "indicator" variables, i.e. to a set of dummy variables. The above command should provide you with column (2) of table 1 in Beyer and Wacker (2022).

References

Araujo, J.T., E. Vostroknutova, M. Brueckner, M. Clavijo and K.M. Wacker (2016): B*eyond Commodities: The Growth Challenge of Latin America and the Caribbean*. Latin American Development Forum Series, Washington, DC.

Beyer, Robert C.M. and Konstantin M. Wacker (2022): “Good Enough for Outstanding Growth: The Experience of Bangladesh in Comparative Perspective.” World Bank Policy Research Paper 10150.

Brueckner, Markus and Daniel Ledermann (2018): “Inequality and economic growth: the role of initial income.” *Journal of Economic Growth* 23(3): 341-366.

Easterly, W., Kremer, M., Pritchett, L., & Summers, L. H. (1993). Good policy or good luck? Country growth performance and temporary shocks. *Journal of Monetary Economics*, 32(3), 459-483.

Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer (2015): "The next generation of the Penn World Table." *American Economic Review* 105 (10): 3150-3182.

Laeven, Luc and Fabian Valencia (2020): “Systemic Banking Crises Database II.” *IMF Economic Review* 68: 307-361.

Moller, L.C. and K.M. Wacker (2017): "Explaining Ethiopia's Growth Acceleration - The Role of Infrastructure and Macroeconomic Policy." *World Development* 96: 198-215

Papageorgiou, Chris, Nikola Spatafora, and Ke Wang (2015): Diversification, Growth, and Volatility in Asia. World Bank Policy Research Working Paper 7380.

Patel, Dev, Justin Sandefur and Arvind Subramanian (2021): “The new era of unconditional convergence.” *Journal of Development Economics* 152: 102687

Pritchett, Lant and Larry Summers (2015): Asiaphoria Meets Regression to the Mean. NBER Working Paper 20573.

Solt, Frederick (2019): The Standardized World Income Inequality Database, Versions 8-9. https://doi.org/10.7910/DVN/LM4OWF, Harvard Dataverse, V6