

Chapter 2

Model Specification and Data

Abstract This chapter outlines the empirical framework and the main variables involved in the analysis of stochastic and deterministic convergence. It then describes the data source used and the exact data series employed to measure real output per worker, real physical capital per worker, human capital, total factor productivity (TFP) and average annual hours worked. As regards the data source, we employ the newest version of the so-called Penn World Table, version 8.0. As measures of real GDP, real physical capital and TFP, we employ the series based on national-accounts data. As for the measure of human capital, we employ the Psacharopoulos (1994) survey of wage equations evaluating the returns to education that transforms average years of schooling data into a human capital index.

Keywords Real GDP per worker · Real physical capital per worker · Human capital · Total factor productivity · Annual hours worked · Penn world table 8.0

After introducing the topic and briefly reviewing the literature on output convergence, we now describe the empirical strategy and data employed in the analysis.

2.1 Model Specification and Definitions of Convergence

The starting point is a standard Cobb-Douglas production function with constant returns to scale on the inputs employed in production, as follows:

$$Y_{it} = A_{it}K_{it}^{\alpha}(h_{it}L_{it})^{1-\alpha} \quad (2.1)$$

where aggregate output Y_{it} is a function of A_{it} which accounts for TFP, the stock of physical capital K_{it} , and human capital augmented labor, which is the product of a human capital index h_{it} (accounting for the amount of human capital per worker) times raw labor given by L_{it} . α is the output elasticity of physical capital, and $1 - \alpha$

is the output elasticity of augmented labor.¹ The production function can be rewritten by expressing aggregate output and physical capital in per worker terms. This renders the following:

$$y_{it} = A_{it} k_{it}^{\alpha} h_{it}^{1-\alpha} \quad (2.2)$$

where y_{it} and k_{it} are output per worker and physical capital per worker, respectively. Applying natural logs to both sides of Eq. 2.2, it renders:

$$\ln y_{it} = \ln A_{it} + \alpha \ln k_{it} + (1 - \alpha) \ln h_{it} \quad (2.3)$$

Equation 2.3 contains the main variables involved in the analysis of convergence conducted below. The aim is to investigate the existence of stochastic and deterministic convergence in output, but also to assess the convergence hypothesis in the main sources of output per worker, i.e. TFP, physical capital per worker, and human capital. This constitutes an important improvement over most of the literature on convergence that has focused only on the study of output convergence and hence neglected the analysis of convergence dynamics in the main sources of output.

Following common practice in the time-series convergence literature, we compute the logarithm of the ratio of country-specific per worker real GDP to the average per worker GDP for the sample of 21 OECD countries. Thus, the variable of interest for unit root testing is relative output levels, i.e. $RI_{it} = \ln(y_{it}/\bar{y}_t)$, where y_{it} represents individual country's real GDP per worker and $\bar{y}_t = (\sum_{i=1}^N y_{it}/N)$ stands for the average real GDP per worker of the group. $i = 1, \dots, N$ stands for the number of countries and $t = 1, \dots, T$ for the time periods. In our case, N equals 21 and T equals 42, thus making a balanced panel composed of 882 observations.

The same normalisation is applied to the stock of physical capital per worker, the human capital index, TFP and average annual hours worked per worker.² Hence, relative physical capital per worker is given by $RK_{it} = \ln(k_{it}/\bar{k}_t)$, where k_{it} represents individual country's real physical capital per worker and $\bar{k}_t = (\sum_{i=1}^N k_{it}/N)$ represents the average real physical capital per worker of the group. For relative human capital, we have $Rh_{it} = \ln(h_{it}/\bar{h}_t)$, with h_{it} representing individual country's human capital per worker and $\bar{h}_t = (\sum_{i=1}^N h_{it}/N)$ the average human capital of the group. As regards relative TFP, $RA_{it} = \ln(A_{it}/\bar{A}_t)$, with A_{it} standing for individual country's TFP and $\bar{A}_t = (\sum_{i=1}^N A_{it}/N)$ being the average TFP of the group. Regarding relative annual hours worked per worker engaged in production, $RH_{it} = \ln(H_{it}/\bar{H}_t)$, where

¹ As shown below, for the computation of the TFP series, Inklaar and Timmer (2013) approximate these output elasticities by assuming perfect competition in factor and good markets. This allows to consider α as the share of GDP not earned by labor.

² We include average annual hours worked in our convergence analysis because the production function could be written in terms of per hour worked rather than in per worker terms.

H_{it} stands for country-specific annual hours per worker and $\bar{H}_t = (\sum_{i=1}^N H_{it}/N)$ is the average annual hours per worker of the group.

By normalising country-specific series of real output per worker or its sources against the average of the respective series, we are able to distinguish country-specific movements from common trends in the respective variable caused by global shocks such as the oil crises of the seventies. However, since this procedure only allows for a very restrictive form of cross-correlation, we employ several methods to allow for general forms of cross-sectional dependence. This comprises the simulation of the bootstrap distribution tailored to the error structure of our panel of each respective series, the use of nonlinear IV or the utilisation of factor models, with the latter allowing for stronger forms of cross-correlation. In a nutshell, a unit root in the log of relative real GDP per worker implies divergence of the series from the average output per worker of the group. By way of contrast, stationarity in the log of relative real GDP per worker levels entails that shocks to real GDP per worker relative to the average affect the series only temporarily, which leads the series to converge after the effect of the shock vanishes. The same reason applies to the relative series constructed for real physical capital per worker, human capital per worker, TFP and average annual hours worked.

As noted by Li and Papell (1999), the concept of stochastic convergence, which implies that the log of relative output per worker is trend stationary, is a weak notion of convergence. This is due to the fact that it allows for permanent differences in per worker output levels across countries through the presence of a linear trend in the deterministic component of the trend function. As a response to that, Li and Papell (1999) propose a stronger definition of convergence, called deterministic convergence, which implies that the log of relative output per worker is mean stationary. For this to hold, it is necessary to eliminate both deterministic and stochastic trends, which would imply that output per worker in one country moves in parallel over the long run relative to average output per worker levels. One can thus infer that deterministic convergence implies stochastic convergence, but not the other way around.

For the sake of completeness, we study both time series definitions of convergence. However, we do not deal in detail with other definitions of convergence using cross-section data. The reasons for this are the following. Firstly, cross-sectional forms of convergence such as conditional β -convergence constitute a much weaker notion of convergence than time series convergence.³ This stems from the fact that cross-section tests are subject to spurious rejections of the null of no convergence

³ β -convergence implies that countries starting from a high income level are expected to exhibit lower income growth than countries beginning with low income levels. The terms conditional and unconditional (absolute) refer to whether convergence takes place after controlling or not for country-specific characteristics, which can account for differences in steady state income levels.

when economies exhibit differing steady states.⁴ Secondly, Ericsson et al. (2001) show that the aggregation of data over several decades may hide convergence. The reverse could also be true, since aggregation may lead to spurious convergence. In addition, cross-section analysis confounds short-run dynamics with long-run features of the data. Thirdly, by taking a panel data approach, we exploit the time series and cross-section dimensions of the data. This allows us to control for conditional convergence through the inclusion of country-specific effects, which proxy for time-invariant compensating differentials among economies. Last but not least, by exploiting the panel dimension of the data, we can combine the transition and steady-state information contained in the cross-section and time-series approaches (Bernard and Durlauf 1996).

2.2 Data Description

Once we have outlined the empirical framework and the main variables involved in our analysis of stochastic and deterministic convergence, we now describe the data source used and the exact data series employed to measure output per worker, physical capital per worker, human capital, TFP and average annual hours worked. As regards the data source, we employ the newest version of the so-called Penn World Table, version 8.0 (henceforth PWT8.0), developed by the joint efforts of Robert Feenstra from the University of California at Davis, and Robert Inklaar and Marcel Timmer from the Groningen Growth and Development Centre at the University of Groningen (see Feenstra et al. 2013a, b, c).⁵

As a measure of aggregate output, we employ a measure of constant-price real GDP, denoted by $RGDP^{NA}$, which represents real GDP at constant 2005 national prices (in million 2005 US\$). This series employs national-accounts growth rates to construct the real GDP series. This series is similar to those of the previous versions of the PWT, though some differences in its computation are pointed out by Feenstra et al. (2013a, b, c). In fact, versions of the PWT prior to 6.0 constructed the real GDP series using a weighted average of the national-accounts growth rates of the GDP components given by private consumption, investment and government consumption. This caused the real GDP growth rate in PWT to differ from the growth rate of real GDP in the national accounts due to these weights. This phenomenon was highly criticised by Johnson et al. (2013) due to the fact that these weights differed across the different versions of PWT. This caveat was addressed by the authors in succeeding versions of PWT by using the national-accounts growth rate of total

⁴ Bernard and Durlauf (1996) further demonstrate that a negative cross-section relationship between initial income and growth is compatible with a class of structural models which violate the time series definition of convergence implied by the equality of long-term forecasts of per capita output for two countries at a fixed time. Along similar lines, Quah (1993) shows that the existence of β -convergence is compatible with a stable cross-section variance in output levels.

⁵ The data are accessible from www.ggd.net/pwt.

GDP rather than that of the C, I and G components. This aggregate output series is comparable across countries and over time.

As regards the stock of physical capital, we employ the real stock of physical capital, denoted by RK^{NA} provided in PWT8.0. As Inklaar and Timmer (2013) point out, there are several clear advantages of using these series versus other physical capital stock series previously developed in the literature. First, this stock of physical capital series accounts for differences in asset composition across countries and over time. Hence, investment is split up into the following categories: structures, transport equipment and machinery, which in turn can be divided into investment in computers, communication equipment, software and other machinery. Thus, this improves over most previous physical capital estimates that assumed total investment in a single homogeneous asset for all countries. Second, an implication from the above is that the depreciation rate of physical capital exhibits variation across countries and over time—instead of assuming a constant depreciation rate—and that the PPP associated with the stocks of physical capital need not be equal to the investment PPP considered in the conventional approach. Third, in computing the initial stock of physical capital, Inklaar and Timmer (2013) replace the restrictive steady-state assumption by considering an initial capital/output ratio. As shown in specification (2.3), we need a measure of total employment to compute both real GDP and real physical capital in per worker terms. For that purpose, we use the employment series in PWT8.0, which tries to measure “the total number of persons engaged in a productive activity within the boundaries of the system of National Accounts. This should include all employees, but also self-employed workers, unpaid family workers that are economically engaged, apprentices and the military” (Inklaar and Timmer 2013, p. 35). Real physical capital per worker represents the stock of physical capital per worker at constant 2005 national prices (in million 2005 US\$).

As far as the measure of human capital is concerned, we employ the human capital index in PWT8.0 obtained on the basis of average years of schooling data for the population aged 15 and over stemming from Barro and Lee (2010), version 1.3 covering the period 1950–2010. They adopt the Psacharopoulos (1994) survey of wage equations evaluating the returns to education to transform these average years of schooling data into a human capital index. In particular, let s_{it} represent the average number of years of education of the adult population in country i at time t and the human capital index be a function of the average number of years of education of the adult population as follows:

$$h_{it} = e^{\phi(s_{it})} \quad (2.4)$$

where h_{it} constitutes an index of human capital per worker. ϕ is a piecewise linear function, with a zero intercept and a slope of 0.134 through the 4th year of education, 0.101 for the next 4 years, and 0.068 for education beyond the 8th year.⁶ Clearly, the rate of return to education (where ϕ is differentiable) is

⁶ See Badunenko and Romero-Ávila (2013, 2014) for other studies employing a similar definition of human capital.

$$\frac{d \ln h_{it}}{ds_{it}} = \phi'(s_{it}) \quad (2.5)$$

As with the other series, the human capital index exhibits cross-country and time-series variability.

As regards the TFP series, we employ the index number provided in PWT8.0, denoted as $RTFP^{NA}$, which is associated with national accounts data and represents TFP at constant national prices (2005 = 1). This series is calculated as follows:

$$RTFP_{i,t-1}^{NA} = \frac{RGDP_{i,t-1}^{NA}}{RGDP_{i,t-1}^{NA}} / Q_{i,t-1}^T \quad (2.6)$$

where $RGDP^{NA}$ is national-accounts based real GDP per worker in PWT8.0 and Q^T is the Törnqvist quantity index of factor inputs (in this case physical capital per worker and human capital).

Since the production function could be expressed in terms of per hour worked instead of in per worker terms, we take advantage of the series of average annual hours worked per worker available in PWT8.0 and also assess whether there has been convergence or not in this measure of raw labour. This measure accounts for the average annual hours worked by persons engaged in productive activity.

In all, we use annual data on the variables described above for 21 OECD countries over the period 1970–2011, for which complete data series were available.⁷ The countries under analysis are Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, the United Kingdom and the United States.

⁷ The time span investigated begins in 1970 because Germany did not have data on employment, TFP and average annual hours worked before that year.