Technical Notes

Country Income Groups

This Report uses the World Bank income classification of 2014 to refer to particular country groups. The classification is based on gross national income per capita and establishes the following four groups: low-income economies (USD 1,045 or less); lower middle-income economies (USD 1,046 to USD 4,125); upper middle-income economies (USD 4,126 to USD 12,736); and high-income economies (USD 12,736 or more).

More information on this classification is available at *http://data.worldbank.org/about/country-classifications*.

Patent Mappings

The case studies in chapters 2 and 3 rely on mappings of patents developed especially for this report. The patent data for these mappings come from the WIPO Statistics Database and the EPO Worldwide Patent Statistical Database (PATSTAT, April 2015). Key methodological elements underlying the mapping exercise include the following:

Unit of analysis

The main unit of analysis is the first filing of a given invention.[134] In consequence, the date of reference for patent counts is the date of first filing. For some historical records – for example, those older than 1930 for USPTO documents – the application date is missing. In such cases, the date of the earliest subsequent filing or the grant date of the first filing has been used. The origin of the invention is attributed to the first applicant of the first filing; whenever this information was missing an imputation strategy has been applied, as described further below.

The only departure from this approach occurs when analyzing the share of patent families requesting protection in each patent office (see figures 2.6, 3.5, 3.10 and 3.14). In this case, an extended patent family definition – known as the INPADOC patent family – has been used instead of the one relying on first filings. In addition, only patent families with at least one granted application have been considered for this analysis, and the date of reference is the earliest filing within the same extended family. The main rationale for using the extended patent family definition and imposing at least one granted patent within the family is to mitigate any underestimation issuing from complex subsequent filing structures, such as continuations and divisionals, and from small patent families of lower quality such as those filed in only one country and either rejected or withdrawn before examination.

Imputing country of origin

When information about the first applicant's country of residence in the first filing was missing, the following sequence was adopted: (i) extract country information from the applicant's address; (ii) extract country information from the applicant's name (see further below); (iii) make use of the information from matched corporations (as described further below); (iv) rely on the most frequent first applicant's country of residence within the same patent family (using the extended patent family definition); (v) rely on the most frequent first inventor's country of residence within the same patent family (again, using the extended patent family definition); and (vi) for some remaining historical records, consider the IP office of first filing as a proxy for origin.

Cleaning applicant names and assigning applicant types

Applicants have been categorized in three broad categories: (a) *Companies*, which includes mostly private companies and corporations, but also state-owned companies; (b) *Academia and public sector*, which includes public and private universities (and their trustees and board of regents), public research organizations, and other government institutions such as ministries, state departments and related entities; (c) *Individuals*, which includes individual first applicants who may or not be affiliated with companies, academia or other entities. A further category, (d) *Not available,* includes all unclassified first applicants.

In order to assign broad type categories to each first applicant, a series of automated steps were performed for each of the six innovation fields underlying the case studies, to clean and harmonize applicant names. The results of this automated process were cross-checked manually – particularly for the top applicants of each type – prompting revision of the strategy and adjustment of parameters in several iterations.

The starting point was the original information about the first applicant's name from the first filing. When this name was missing, the most frequent first applicant's

134. Mappings include data on utility models whenever available.