ILO modelled estimates and projections: Data considerations and methodological approach

Data considerations

The source of estimates and projections of employment, unemployment, status in employment, employment by economic activity and working poverty is the ILO's Trends Econometric Models (TEM). The ILO Research Department and ILO Department of Statistics designed and actively maintain several econometric models, which are used to produce estimates of labour market indicators in the countries and years for which country-reported data are unavailable. These allow the ILO to produce and analyse global and regional estimates of key labour market indicators and related trends.

The TEM is used to produce estimates and projections – disaggregated by age and sex as appropriate – of unemployment, employment, employment by industry and broad occupational group, status in employment and labour productivity (output per worker). The output of the model is a complete matrix of data for 188 countries. The country-level data can then be aggregated to produce regional and global estimates of labour market indicators, such as the unemployment rate and the employment-to-population ratio.

The purpose of estimating labour market indicators for countries with missing data is to obtain a balanced panel dataset in order to compute, every year, regional and global aggregates with consistent country coverage. Despite the use of all available information, estimates for countries with very limited labour market information remain highly uncertain. Therefore, the estimation methodology ensures that changes in global and regional aggregates are mostly driven by countries with real information. This also means that estimates of labour market indicators for countries with limited nationally reported labour market information should not be considered as "true" data, and great care needs to be applied as to its usability in further analysis, especially at the country level.

Methodological approach

Prior to running the TEM, labour market information specialists in the Research Department, the Department of Statistics and ILO field offices, evaluate existing country-reported data and select only those observations deemed sufficiently comparable across countries using criteria including: (1) type of data source; (2) geographic coverage; and (3) age group coverage.

- With regard to the first criterion, in order for data to be included in the model, they must be derived from either a labour force survey, other sufficiently comparable household survey, or a population census. National labour force surveys are generally similar across countries, and the data derived from these surveys are more readily comparable than data obtained from other sources. A strict preference is therefore given to labour force survey-based data in the selection process. However, many developing countries which lack the resources to carry out a labour force survey do report labour market information based on other types of household surveys or population censuses. Consequently, due to the need to balance the competing goals of data comparability and data coverage, some other survey-based estimates and population census-based data are included in the model.

- The second criterion is that only nationally representative (i.e. not prohibitively geographically limited) labour market indicators are included. Observations which correspond to only urban or only rural areas are not included, as large differences typically exist between rural and urban labour markets, and using only rural or urban data would not be consistent with benchmark data such as GDP.

- The third criterion is that the age groups covered by the observed data must be sufficiently comparable across countries. Countries report labour market information for a variety of age

groups and the age group selected can have an influence on the observed value of a given labour market indicator.

Apart from country-reported labour market information, the TEM uses the following benchmark files:

- United Nations World Population Prospects for population estimates and projections;
- ILO Labour Force Estimates and Projections (LFEP) for labour force estimates and projections;
- IMF/World Bank data on GDP (PPP, per capita GDP and GDP growth rates) from the World Development Indicators and the World Economic Outlook database;
- Oxford Economics quarterly national accounts forecasts for 41 countries with short-term forecasts
- World Bank poverty estimates from the PovcalNet database.

Estimates of labour market indicators

The TEM produce estimates of unemployment rates to fill in missing values in the countries and years for which country-reported data are unavailable. Multivariate regressions are run separately for different regions in the world in which unemployment rates, broken down by age and sex (youth male, youth female, adult male, adult female), are regressed on GDP growth rates. Weights are used in the regressions to correct for biases that may result from the fact that countries that report unemployment rates tend to differ (in statistically important respects) from countries that do not report unemployment rates.[1] For the current year, a preliminary estimate is produced, using quarterly and monthly information available up to the time of production of the estimates. The model also estimates employment by status using similar techniques to impute missing values at the country level. In addition to GDP growth rates, the variables used as explanatory variables include: the value added shares of the three broad sectors in GDP, per capita GDP and the share of people living in urban areas. Additional econometric models are used to produce global and regional estimates of working poverty and employment by economic class.[2]

Projections of labour market indicators

Unemployment rate projections are obtained using the historical relationship between unemployment rates and GDP growth during the worst crisis/downturn period for each country between 1991 and 2005, and during the corresponding recovery period.[3] This is done through the inclusion of interaction terms of crisis and recovery dummy variables with GDP growth in fixed effects panel regressions.[44] Specifically, the logistically transformed unemployment rate is regressed on a set of covariates, including the lagged unemployment rate, the GDP growth rate, the lagged GDP growth rate and a set of covariates consisting of the interaction of the crisis dummy and the interaction of the recovery-year dummy with each of the other variables.

---

[1] For instance, if simple averages of unemployment rates in reporting countries in a given region are used to estimate the unemployment rate in that region, and the countries that do not report unemployment rates differ from reporting countries with respect to unemployment rates, without such a correction mechanism the resulting estimated regional unemployment rate would be biased. The "weighted least squares" approach adopted in the TEM corrects for this potential problem.

[2] See Kapsos and Bourmpoula, 2013. Available at: http://www.ilo.org/wcmsp5/groups/public/---dgreports/---inst/documents/publication/wcms_216451.pdf

[3] The crisis period comprises the span between the year in which a country experienced the largest drop in GDP growth and the "turning point year" when growth reached its lowest level following the crisis before starting recover. The recovery period comprises the years between the "turning point year" and the year when growth has returned to its pre-crisis level.

[4] In order to project unemployment during a country's recovery period, the crisis-year and recovery-year dummies were adjusted, based on the following definition: a country was considered to be "currently in crisis" if the drop in GDP growth after 2007 was larger than 75 per cent of the absolute value of the standard deviation of GDP growth over the 1991–2008 period and/or larger than 3 percentage points.

Separate panel regressions are run across three different groupings of countries, basedon:

(1) geographic proximity and economic/institutional similarities;

(2) income levels;[5]

(3) level of export dependence (measured as exports as a percentage of GDP).[6]

The rationale behind these groupings is as follows: Countries within the same geographic area or with similar economic/institutional characteristics are likely to be similarly affected by a crisis and have similar mechanisms to attenuate the impact of the crisis on their labour markets. Furthermore, because countries within given geographic areas often have strong trade and financial linkages, the crisis is likely to spill over from one country to its neighbour (e.g. Canada's economy and labour market developments are intricately linked to developments in the United States). Countries with similar income levels are also likely to have similar labour market institutions (e.g. social protection measures) and similar capacities to implement fiscal stimulus and other policies to counter the crisis impact. Finally, as a decline in exports is often a key crisis transmission channel from developed to developing countries, countries are grouped according to their level of exposure to this channel, as measured by their exports as a percentage of GDP. The impact of the crisis on labour markets through the export channel also depends on the type of exports (the affected sectors of the economy) involved, the share of domestic value added in exports and the relative importance of domestic consumption (for instance, countries such as India and Indonesia, with large domestic markets, tend to be less vulnerable to external shocks than countries such as Singapore and Thailand). These characteristics are controlled for by using fixed effects in the regressions. In addition to the panel regressions, country-level regressions are run for countries with sufficient data. The ordinary least squares country-level regressions include the same variables as the panel regressions.

To take into account the uncertainty surrounding GDP prospects, as well as the complexity of capturing the relationship between GDP and unemployment rates for all the countries, a variety of ten (similar) multilevel mixed-effects linear regressions (varying-intercept and varying-coefficient models) are utilized. The main component that changes across these ten versions is the lag structure of the independent variables. The potential superiority of these models lies in the fact that not only is the panel structure fully exploited (e.g. increased degrees of freedom), but it is also possible to estimate the coefficients specifically for each unit (country), taking into account unobserved heterogeneity at the cluster level and correcting for the potential problem that the independent variables are not correlated with the random effects term. A final projection is generated as a simple average of the estimates obtained from the three group panel regressions and also, for countries with sufficient data, the country-level regressions.

Youth labour market indicators

Labour market indicators for the sub-populations youth-female, youth-male, adult-female and adult-male have been estimated using the same regression techniques as the aggregate indicators. However, the estimates are adjusted using the shares in the population implied by the labour force survey estimates so that the implied sum of the sub-populations equals the aggregate rate. This means that country data on sub-populations could differ from reported rates in other sources when the underlying shares of the sub-population in the labour force differ from the ILO's estimates.

[5] The income groups correspond to the World Bank income group classification of four income categories, based on countries' (GNI) per capita.

[6] The export dependence-based groups are: highest exports (exports ≥ 70 per cent of GDP); high exports (exports < 70 per cent but ≥ 50 per cent of GDP); medium exports (exports < 50 per cent but ≥ 20 per cent of GDP); and low exports (exports <20 per cent of GDP).

Short-term projection model

For a subset of countries, the preliminary unemployment estimate for the current year and the projection for the following year are based on results from a country-specific short-term projection model. The ILO maintains a database on monthly and quarterly unemployment flows that contains information on inflow and outflow rates of unemployment, estimated on the basis of unemployment by duration. A multitude of models are specified that either project the unemployment rate directly or determine both inflow and outflow rates, using ARIMA, VARX and combined forecast techniques. All estimated models are evaluated on an eight-quarter ahead rolling pseudo out-of-sample forecasting evaluation starting in Q1 2009, among which five models are selected using a weighting of the mean and maximum forecast error. The top five model forecasts are then averaged.