HERMREG: A regionalisation model for Belgium

April, 2007

Hoorelbeke, D*[c]., Lohest, O.[d], Bassilière, D.[a], Bossier, F[a], Bracke, I[a] and Caruso, F[b].

a: Federal Planning Bureau
b: Institut Bruxellois de Statistique et d'Analyse (IBSA)
c : Research Center of the Flemish Government
d: Institut Wallon de l'évaluation, de la prospective et de la statistique (IWEPS)
* : corresponding author, dh@plan.be

Abstract:

HERMREG is a first (and partial) attempt to regionalise HERMES. HERMES is a multisectoral macroeconomic model of the Belgian economy, describing 16 industrial sectors. HERMES, however, is only about national variables.
HERMREG is developed to regionalise the national projections from HERMES, in order to have economic projections for the three Belgian regions (Brussels, Flanders and Wallonia). In this first version, the main focus is on four key variables: value added, investments, employment and wages. The model also contains an environmental module. The regionalisation method is a combined shift-share regression method. Equations are estimated for each region (3) and sector (13), using a sample from 1980 to 2003/4.

Keywords: regional model, econometric model

1. Introduction

Belgium is a federal country, which is divided in three regions: the Brussels Capital Region, the Flemish Region and the Walloon Region[1]. These regions have their own government and political institutions. The regions deal with territory related matters, such as agriculture, environment, employment, energy, economy, foreign trade, … . During the past decades, more and more political power is transferred from the national level towards the regions. To this end, a regional economic model can support the decision making process at the regional level. Therefore, the three regions and the (national) Federal Planning Bureau started a project to develop a multiregional economic model named HERMREG. The Federal Planning Bureau already manages an economic model of Belgium (at the national level): HERMES (which is an acronym for Harmonised Econometric Research for Modelling Economic Systems). HERMES, as can also be deduced from its full name, is also used in other European countries to produce economic projections and to implement simulations. HERMES models exist in the following countries: The Netherlands, France, Germany, Ireland, Italy, the United Kingdom and Belgium[2].

HERMES is mainly used to produce medium term economic forecasts (5-6 years ahead) and to analyse the economic impact of policy measures and external shocks, both are however restricted to the national level. The model is mainly demand orientated, although also supply aspects play a role. HERMES is composed of 14 branches (agriculture, contruction, …), 15 consumption goods and 5 institutional sectors (households, firms, …). HERMES is an econometric model[3]: the model is based on a theoretical model, from which equations are derived that are to be estimated via time series.

In HERMES, however, the regions are not separately modelled. As a result, there are no regional projections nor regional simulation studies available. To meet this shortcoming HERMREG is developped. The first version of HERMREG is mainly focussed on four variables: value added, employment, investments and salaries. These variables are modelled on a sectoral and

---

[1] Besides the division in regions, Belgium is also decomposed in three communities (the Flemish Community, the French Community and the German Community). These communities deal with persons related matters (such as cultural affairs, education, health, well-being and language usage).
[2] See Commission of the European Communities (1993).
[3] For more details, see Bossier et al (2000).

regional level using a top-down method. Regional projections of the following variables are available through HERMREG: gross domestic product, value added, employment (salaried and self-employment), net commuting flows, investments, wages, productivity and unemployment. These variables, except for GDP, unemployment and net commuting flows, are available for 13 branches.

This paper mainly discusses the methodology used for the construction of this first version of HERMREG. The methodology is based on two techniques: shift-share analysis and time series regressions (see Section 2). HERMREG produces regional weights, which are then applied on the national sectoral forecasts to obtain the regional projections. These weights are calculated using the combined shift-share and regression methods. A similar shift-share regression method is also used in REGINA (Koops and Muskens, 2005). REGINA is a Dutch multiregional economic model.

This paper is structured as follows. Section 2 discusses in detail the applied methodology. Section 3 concludes.

2. Methodology

2.1. Basic principles of a shift-share analysis

A shift-share analysis is used to decompose the growth of a certain variable, say Y, in branch i from region j. The increase of $Y_{ij,t}$, i.e. $\Delta Y_{ij,t}$, is decomposed in three components[4]:

- a national growth factor (or share effect), $n_{ij,t}$;
- a industry-mix factor (or proportional shift), $s_{ij,t}$;
- a regional effect (or differential shift), $r_{ij,t}$.

Before calculating these three components, we first introduce some notation:

- increase of Y in branch i from region j: $\Delta Y_{ij,t} = Y_{ij,t} - Y_{ij,t-1}$;
- national growth of Y: $y_t = (Y_t - Y_{t-1})/Y_t$ with $Y_t = \sum_i \sum_j Y_{ij,t}$;
- national sectoral growth of Y: $y_{i.,t} = (Y_{i.,t} - Y_{i.,t-1})/Y_{i.,t}$ with $Y_{i.,t} = \sum_j Y_{ij,t}$;
- regional sectoral growth of Y: $y_{ij,t} = (Y_{ij,t} - Y_{ij,t-1})/Y_{ij,t}$.

The national growth factor, $n_{ij,t}$, is computed as follows,

(1) $$n_{ij,t} = y_t Y_{ij,t-1}.$$

The second part of the decomposition is the industry-mix component, $s_{ij,t}$, which is calculated as:

(2) $$s_{ij,t} = (y_{i.,t} - y_t) Y_{ij,t-1}.$$

The regional effect, $r_{ij,t}$, equals

(3) $$r_{ij,t} = (y_{ij,t} - y_{i.,t}) Y_{ij,t-1}.$$

---

[4] In this note, differences are between *t* and *t-1*, since this seems most interesting for HERMREG. This is not necessary, however. In shift-share analysis one sometimes takes longer time periods to compute differences.

The differential shift component is that part of a region's growth which remains unexplained, and which can be attributed to region specific characteristics (or other explanatory variables). It is obvious that these three effects sum up to the difference in the regional variable, $\Delta Y_{ij,t}$, i.e.

$$\text{(4)} \qquad \Delta Y_{ij,t} = n_{ij,t} + s_{ij,t} + r_{ij,t}.$$

The presentation above is in absolute figures. The shift-share analysis can also be done in relative figures (i.e. growth numbers) by dividing the equations (1)-(4) by $Y_{ij,t-1}$.

It is also possible to represent a region's growth (in Y) through these components, but now on a regional level, instead of the regional sectoral level. That is, the different components are summed over the branches i, to obtain (in growth figures):

$n_{.j,t} = \sum_i n_{ij,t} / \sum_i Y_{ij,t-1}$ (= y_t),

$s_{.j,t} = \sum_i s_{ij,t} / \sum_i Y_{ij,t-1}$, and

$r_{.j,t} = \sum_i r_{ij,t} / \sum_i Y_{ij,t-1}$.

So, a region's growth differs from national growth by an industry-mix factor and a regional factor:

$$\text{(5)} \qquad y_{.j,t} = y_t + s_{.j,t} + r_{.j,t} \;\text{ where }\; y_{.j,t} = (Y_{.j,t} - Y_{.j,t-1}) / Y_{.j,t-1} \;\text{ and }\; Y_{.j,t} = \sum_i Y_{ij,t}.$$

The industry-mix factor is positive if the region has a relative high presence of high growth industries: $s_{.j,t}$ is positive if the weighted sum of sectoral growth rates in the region exceeds the national growth rate, where the weights are the shares of the branches i in the total region's j Y in the time period $t-1$, i.e. the weights are $Y_{ij,t-1} / Y_{.j,t-1}$.

2.2. The shift-share and regression method of REGINA

In REGINA (Koops and Muskens, 2005), only one variable is modelled by means of this combined shift-share-regression method, namely employment. REGINA is only about employment, no other variables are regionalised in REGINA.

The dependent variable in the regression part is not the differential shift $r_{ij,t}$ itself, but a closely related variable:

$$\frac{y_{ij,t}}{y_{i.,t}},$$

where in REGINA y equals employment growth. Remind that the differential shift equals (expressed in growth figures):

$$r_{ij,t} = y_{ij,t} - y_{i.,t}. \tag{3'}$$

So, instead of using a difference indicator, REGINA uses the differential ratio as dependent variable in the regression analysis. The explanatory variables are the following:

- *labour supply*: population growth in the 15-64 age group, potential labour force (in absolute terms), and population density;
- *agglomeration effects*: a variable measuring the proximity of concentrations of employment;
- *accessibility*: accessibility indicators with and without congestion;
- *proximity of economic mainports*: distance indicators to Schiphol Airport, Port of Rotterdam, Ruhrgebiet, Brussels, Antwerp;
- *sector representation*: a sector specialization variable and an economic diversity variable (Hirshman-Herfindahl index);
- *other factors*: intensity of land use, suburbanization and a dummy variable for Flevoland.

The regressions are done for given "sector" i, with the different regions j as observation points. Firstly, the shift-share analysis is done for three time periods: 1970-1983, 1983-1993 and 1993-2000. There are 40 regions (NUTS3) [5], which gives 40 x 3 observations for each of the six "sectors". The "sectors" are: manufacturing, trade, transport and communication, commercial services, government, and non-profit. The authors argue that the employment growth in the agriculture "sector" is fully explained by the industry-mix component, and therefore its differential part is not analysed by a regression. How the parameters are exactly estimated is not entirely clear from the article, but supposedly just by pooling the observations and then applying OLS.

[5] In a second step, the analysis is done at an even more disaggregated level, namely for the 496 municipalities of the Netherlands.

In the six regressions the following variables are found to be significant at 5% (in parentheses the number of times the variable is significant, its sign, and, sometimes, the "sector(s)" for which the variable is significant):

- population growth (6, +),
- labour force (1, +, commercial services),
- population density (2, trade (-) and government (+)),
- "sector" specialization (6, -),
- national accessibility (2, +, trade and commercial services),
- congestion (1, +, trade),
- international accessibility (4, + and -),
- intensity of land use (1, -, manufacturing),
- dummy for Flevoland (1, non-profit).

REGINA is then used to make projections of employment growth over the period 2002-2010, so not for every year separately. The primary inputs for these projections are the projections of the CPB with respect to national employment growth (for the industry mix part) and regional population projections ("PRIMOS bevolkingsprognoses"). The other explanatory variables are assumed to remain unchanged.

2.3. The mixed shift-share-regression method of HERMREG

The shift-share analysis presented above is adapted such that there are only two components: an industry-mix component and a differential shift. These are then redefined as follows:

(2') $$s_{ij,t} = y_{i..,t} Y_{ij,t-1}, \text{ and}$$

(3) $$r_{ij,t} = (y_{ij,t} - y_{i..,t}) Y_{ij,t-1},$$

or, alternatively, in growth numbers:

(2'') $$s_{ij,t} = y_{i..,t}, \text{ and}$$

(3') $$r_{ij,t} = y_{ij,t} - y_{i..,t}.$$

By using this shift-share analysis, the first part of the regional growth projections, the industry-mix factor, comes from the national model HERMES. The differential shift is to be modelled in a regression framework. The differential shift is about the difference between regional sectoral growth and national sectoral growth, and could be explained by regional characteristics.

The presentation beneath assumes that the shift-share components are expressed in growth numbers, instead of absolute changes.

Suppose $X_{ij,t}$ is a vector containing different explanatory variables, which are used to model the differential shift:

(6) $$r_{ij,t} = X_{ij,t}\beta_{ij} + \varepsilon_{ij,t} \quad \forall j, \forall i$$

So, by using data from the past (*t=1,…T)*, a regression equation is estimated, resulting in:

(7) $$r_{ij,t} = X_{ij,t}\hat{\beta}_{ij} + \hat{\varepsilon}_{ij,t} \quad \forall j, \forall i$$

How are projections made for $Y_{ij,T+k}$ ($k>0$), in this shift-share oriented method? The industry-mix part is easy, since these are merely the national sectoral growth projections produced by HERMES, say $\bar{y}_{i\cdot,T+k}$. Given projections for the explanatory variables $\hat{X}_{ij,T+k}$, the differential shift can also be forecasted:

(8) $$\hat{r}_{ij,T+k} = \hat{X}_{ij,T+k}\hat{\beta}_{ij}\,.$$

Summarising, the forecast of regional sectoral growth for branch *i* in region *j* for time period *T+k* equals:

(9) $$\hat{y}_{ij,T+k} = \bar{y}_{i.,T+1} + \hat{r}_{ij,T+k}\,.$$

It is, however, very unlikely that this projection will be consistent with the national sectoral HERMES projection, that is in general it holds, e.g. for $k = 1$, that

(10) $$\sum_j w_{ij,t}\,\hat{y}_{ij,t+1} \neq \bar{y}_{i.,t+1}, \text{ where } w_{ij,t} = Y_{ij,t} / Y_{i.,t}\,.$$

Or stated otherwise: the initial regional projections will not sum up to the national forecasts made by HERMES. For (10) to be an equality, the shift-share-regression projections (or alternatively, the HERMES projections) should be adapted. There are several ways to do this, but in HERMREG the following method is applied. The initial growth projections of HERMREG, $\hat{y}_{ij,t+k}$, serve to calculate (provisional) values of the variable in level, i.e.

(11) $$\hat{Y}_{ij,t+1} = (1 + \hat{y}_{ij,t+1}) Y_{ij,t} \quad \text{and,}$$

(12) $$\hat{Y}_{ij,t+k} = (1 + \hat{y}_{ij,t+k}) \hat{Y}_{ij,t+k-1} \quad \text{for } k > 1.$$

The following step consists of using these level projections to calculate the weight of each region *j* in a certain branch *i*,

(13) $$\hat{w}_{ij,t+k} = \frac{\hat{Y}_{ij,t+k}}{\hat{Y}_{i.,t+k}} \quad \text{with} \quad \sum_j \hat{w}_{ij,t+k} = 1.$$

Ultimately, the final step consists of applying these weights to the national sectoral projections of HERMES, $\overline{Y}_{i.,t+k}$, to compute the final regional sectoral projections (in level):

(14) $$\tilde{Y}_{ij,t+k} = \hat{w}_{ij,t+k} \overline{Y}_{i.,t+k}.$$

Thus, in short, HERMREG is used to compute endogenous weights to regionalise the national HERMES projections. Now, the regional projections, $\tilde{Y}_{ij,t+k}$, sum up to their national counterpart, $\overline{Y}_{i.,t+k}$ (or equivalently, equation (10) is now an equality).

2.4. Econometric methodology

Before the estimation procedure, all variables are tested for unit roots, initially by means of the augmented Dickey-Fuller test, and in case of doubt, also the Phillips-Perron test is implemented[6]. Non-stationary variables are used in first difference.

For the different variables of interest equation (6) is estimated using OLS, equation by equation. At this moment, there is no correction for the (likely)

[6] All test results are available from the authors upon request.

presence of contemporaneous correlation between the error terms (of the different equations). At least a SUR estimation (Seemingly Unrelated Regressions proposed by Zellner, 1962) seems possible. The SUR estimator is in fact a feasible GLS (Generalised Least Squares) estimator, which makes use of an estimate of the covariance matrix of the residuals.

The final equation for each variable is based on the following selection process. First of all, explanatory variables are only kept in the equation if they are significant at the 10%-level[7]. To choose between different versions of equations, the (centred) R^2 and the Theil coefficient are used. The Theil coefficient is a measure of fit, which in fact equals a rescaled version of the root mean squared error:

$$\text{(15)} \qquad \text{Theil coef.} = \frac{\sqrt{\sum_{t=1}^{T}(\hat{y}_t - y_t)^2 / T}}{\sqrt{\sum_{t=1}^{T}\hat{y}_t^2 / T} + \sqrt{\sum_{t=1}^{T} y_t^2 / T}}.$$

In the numerator one finds the (in sample) root mean squared error. The denominator serves to scale the root mean squared error, such that the result always lies between 0 and 1. If the Theil coefficient equals 0, one has a perfect fit; the closer to one, the worse the fit.

All equations are tested for serial correlation using the Lagrange Multiplier test and the Ljung-Box Q-test. They are also tested for heteroskedasticity by White's heteroskedasticity test. If necessary, corrections are made[8].

Rather than basing the estimation on a theoretical model, the approach (to select the explanatory variables) is ad hoc and exploratory. The selection method can, to a large extent, be described as a general-to-specific modelling strategy (see e.g. Hendry (2000), although his research is more concerned with *automated* general-to-specific modelling strategies), i.e. one starts with an equation including all possible explanatory variables, from which one then eliminates the insignificant variables.

[7] Although sometimes, on an ad hoc base, variables with p-values slightly higher than 10% are also kept in the equation.

[8] All test results are available from the authors upon request.

2.5 Set of explanatory variables

What follows is a short overview of the variables which are used in the regressions to model the differential growth rates of value added, investments, employment and wages. Of course, per equation only a small subset of this set is actually used.

- a constant,
- a dummy,
- lags of the differential growth of value added: DIFFj_QVOi(t-k),
- lags of the regional value added growth rate in branch i: GROj_QVOi(t-k),
- lags of the national growth rate of value added in branch i: GROT_QVOi(t-k),
- lags of the national exports growth rate in branch i: GRO_QXOi(t-k),
- lags of the national productivity growth rate in branch i: GRO_PRODT_QVOi(t-k),
- lags of the regional productivity growth, GROj_PRODi(t-k),
- lags of the differential growth rate of investments in branch i: DIFFj_IOi(t-k),
- lags of the regional investments growth rate in branch i: GROj_IOi(t-k),
- lags of the national investments growth rate in branch i: GROT_IOi(t-k),
- lags of the index of regional investment specialisation in branch i: HINDEXij(t-k),
- lags of the regional growth rate of the (active) population: GROj_NPO(t-k) and GROj_NPA(t-k),
- the real long interest rate, RLBE(t);
- lags of the national capacity utilisation rate of branch i: QRi(t-k),
- lags of the differential growth of wage per employee, DIFFj_WBNFi(t-k),
- lags of the regional growth rate of wages (per employee) in sector i: GROj_WBNFi(t-k),
- lags of the national growth of wage per employee, GROT_WBNFi(t-k),
- lags of the differential salaried employment growth, DIFFj_NFi(t-k),
- lags of the regional employment growth, GROj_NFi(t-k),

- lags of the national growth rate of salaried employment in sector i: GROT_NFi(t-k),

where i stands for the region and j for the sector.

To model the differential growth rate of value added the most prevalent explanatory variables are lags of the dependent variable, the national growth of value added (per sector) and national export growth (per sector).The measure of fit, R2, is for these equations larger than 70%, and also the Theil coefficient is not too high. Thus, both measures indicate a good fit. In general, we obtain a reasonable fit for the value added equations: the (unweighted) average R2 is about 0.75 for both Brussels and Flanders, and about 0.73 for Wallonia.

Some general observations on the investment equations: first, the regressions yield very heterogeneous results across regions and branches. This indicates that the determinants of the differential growth rate are specific to regions and branches. Second, the estimations show that past values of the dependent variable are significant in a large number of equations, implying a very high degree of persistence of the underlying time series. Third, in more than 60 percent of the equations a national aggregate (the investment growth rate or value added growth rate) appears to be significant. Fourth, the cost of capital (RLBE) and capacity of utilization (QRi) appear statistically insignificant in almost all equations.

The variables which occur most in the equations for the differential growth of the wage per employee are the first and second lag of the independent variable itself (DIFFj_WBNFi(t-1) and DIFFj_WBNFi(t-2)), and the first lag of the national growth of the wage per employee (GROT_WBNFi(t-1)). Taking all lags of a variable together, the most used variables are: the differential growth of wages per employee (DIFFj_WBNFi), the differential growth of employment (DIFFj_NFi), the national growth of wages per employee (GROT_WBNFi) and the national productivity growth (GRO_PRODT_QVOi).

The two most frequent explanatory variables are the lags of the regional differential growth of salaried employment and the differential growth (or also the national growth rate) of value added. It is also interesting to note that the wage (per employee) is more often a significant determinant of the differential employment growth in Flanders and Wallonia than in Brussels.

The variables which occur most in the equations concerning the differential growth rate of self-employment are the first lag of the differential growth rate of total employment (DIFFj_NFi(t-1)), the second lag of the differential growth rate of wages (DIFFj_WBNFi(t-2)), and the first lag of the growth of working age population (GROj_NPA(t-1)). We found also that the regional or national level of unemployment rate has a statistically significant impact on the differential shift of self-employment.

Commuting is an important economic phenomenon in Belgium. Commuting is interesting because it acts as a form of quasi-mobility of labour (people travel long or short distances rather than moving). In this way, commuting helps to facilitate regional adjustment to asymmetric shocks (such as an idiosyncratic fall in the demand for a region's product). It can also play an important role in reducing unemployment and income disparities between regions.

To take into account commuting in HERMREG, three equations (one per region) for net commuting flows are estimated. The net commuting flow of a region is defined as the number of people of that region working in the two other regions minus the number of people of the two other regions working in the region. The commuting decision is modelled as depending on regional economic conditions in the destination regions, on regional economic conditions in the source region and by the past value of the net commuting flows.

3. Conclusion

HERMREG is developped in order to facilitate the decision making proces at the regional level. This econometric model uses a mixed shift-share regression method. This method permits to develop reasonably fast a tool to produce regional projections, which are coherent with the national projections.

Reference List

Commission of the European Communities, 1993. "HERMES: Harmonised econometric research for modeling economic systems", Elsevier, Amsterdam, 720 p.

Bossier F., Bracke I., Stockman P. and F. Vanhorebeek (2000). "A description of the HERMES II model for Belgium", Federal Planning Bureau WP 5.00.

Koops, O. and J. Muskens (2005), "REGINA. A model of economic growth prospects for Dutch regions", in: F. van Oort, M. Thissen and L. van Wissen, A survey of spatial economic planning models in the Netherlands, Den Haag: Netherlands Institute for Spatial Research, 104-115.