Appendix

Appendix 1. Decomposition of changes in value added

This report provides a decomposition of deviation in value added per capita in the region from the national average value added for Poland into industry and services, where each of these sections includes a further division into employment (labour input), productivity and residual. For the economy we explain the value added index I_{VA}, which can be expressed as follows:

$$I_{VA} = \frac{VAP_i}{VAP_K} = \frac{VA_i}{P_i}\frac{P_K}{VA_K} = \frac{ER_i P_i W_i}{P_i}\frac{P_K}{ER_K P_K W_K} = \frac{ER_i}{ER_K}\frac{W_i}{W_K}$$

where:

i – examined region
K – country where the region is located
VAP – value added per capita
VA – value added
P – population
ER – employment rate (in the overall population)
W – productivity (VA per worker)

In view of the above, the index value is simply a product of quotients of employment rates and productivity indicators. The result can be expressed as follows:

$$I_{VA} = \frac{ER_i}{ER_K} + \frac{W_i}{W_K} + R_i$$

$$R_i = \frac{ER_i}{ER_K}\frac{W_i}{W_K} - \left(\frac{ER_i}{ER_K} + \frac{W_i}{W_K}\right)$$

The above decomposition can be applied in analyses for the economy as a whole, without any division into sectors. For sector decomposition, contributions of particular sectors need to be additionally weighted by the share of value added of a given sector s in the total national value added. Thus, the equation takes the following form:

$$I_{VA} = \sum_s \left(\frac{VA_{sK}\dfrac{ER_{si}}{ER_{sK}} - VA_{sK}}{\sum_s VA_{sK}} + \frac{VA_{sK}\dfrac{W_{si}}{W_{sK}} - VA_{sK}}{\sum_s VA_{sK}} + \frac{VA_{sK}R_{si} - VA_{sK}}{\sum_s VA_{sK}} \right)$$

In this report, the analysis is limited to two sectors – industry and services – due to the lack of reliable time series on employment in agriculture at the subregional level.

Table A.1.

Change in productivity and employment in subregions in 2005 compared with 1995 (1995=100)

	employment			productivity		
	total	industry	services	total	industry	services
bialskopodlaski	81.9	80.4	82.5	130.2	121.7	132.1
białostocko-suwalski	97.0	91.2	100.3	141.8	130.8	144.6
bielsko-bialski	72.2	58.0	87.7	175.4	187.4	162.7
bydgoski	89.4	76.2	101.1	141.8	133.6	142.8
centralny śląski	72.4	55.3	95.1	172.8	188.2	153.0
chełmsko-zamojski	90.7	94.7	88.8	129.0	111.9	136.8
ciechanowsko-płocki	95.7	78.5	110.7	176.7	230.0	143.7
częstochowski	88.6	83.6	93.7	162.1	177.9	150.0
elbląski	89.0	88.6	89.3	153.6	130.3	169.5
ełcki	89.4	90.1	89.0	152.2	139.3	157.5
Gdańsk-Gdynia-Sopot	81.7	60.4	94.5	168.1	166.4	166.4
gdański	94.3	94.8	93.9	164.3	160.4	167.0
gorzowski	84.2	75.6	90.8	146.5	165.5	134.9
jeleniogórsko-wałbrzyski	76.8	66.7	87.1	165.1	160.5	160.7
kaliski	93.1	93.7	92.5	165.2	155.4	172.2
koniński	86.9	81.9	92.4	150.4	132.2	163.0
koszaliński	78.9	76.5	80.3	179.7	155.3	188.9
krakowsko-tarnowski	83.2	70.3	96.9	169.3	171.8	160.4
Kraków	87.5	59.2	106.3	174.6	192.6	164.5
krośnieńsko-przemyski	84.1	71.8	93.7	134.1	126.2	134.1
legnicki	82.7	74.2	92.2	164.8	179.9	151.4
lubelski	91.8	77.2	101.4	156.2	144.4	159.7
łomżyński	83.5	76.3	87.5	172.8	197.3	161.9
łódzki	91.8	80.0	102.8	149.5	190.0	126.8
Łódź	78.5	58.3	92.7	173.3	186.9	161.8
nowosądecki	88.2	81.9	92.2	147.2	130.1	154.0
olsztyński	91.6	85.1	95.6	157.3	145.9	161.2
opolski	76.6	62.9	89.9	164.1	186.2	147.0
ostrołęcko-siedlecki	85.0	75.5	90.8	155.8	217.2	135.7
pilski	92.9	101.1	86.4	158.0	160.4	160.9
piotrkowsko-skierniewicki	81.5	72.2	91.8	176.1	172.5	176.0
Poznań	97.2	74.3	111.5	168.8	160.0	172.3
poznański	106.6	105.8	107.6	170.5	157.7	180.3
radomski	94.7	64.8	120.7	155.0	217.2	124.1
rybnicko-jastrzębski	72.4	59.1	92.9	180.7	211.2	148.0
rzeszowsko-tarnobrzeski	90.6	79.2	101.8	172.0	172.0	168.0
słupski	89.2	85.8	91.6	133.4	150.2	126.0
szczeciński	80.2	64.9	92.7	152.9	148.0	145.9
świętokrzyski	77.3	66.5	84.3	181.1	187.2	176.8
toruńsko-włocławski	88.0	80.0	94.3	140.2	135.6	139.2
Warszawa	102.1	67.6	116.4	178.6	167.4	177.1
warszawski	103.7	77.2	128.6	167.4	163.5	166.5
Wrocław	85.8	57.6	102.3	159.5	175.2	149.5
wrocławski	99.9	100.2	99.7	176.0	184.3	170.8
zielonogórski	86.1	76.1	94.0	154.3	202.1	131.6
Poland	**86.6**	**71.7**	**98.8**	**164.6**	**166.9**	**158.5**
Poland excluding Warsaw	85.2	71.9	96.7	161.9	167.0	154.7
Poland excluding cities	87.2	77.1	95.8	158.9	163.1	153.0

Below-average changes are presented in grey colour.

Source: Own calculation based on CSO regional accounts.

Table A.2.
Decomposition of sources of variation in value added per inhabitant from national average - 1995

		industry		services		
	total	employment	productivity	employment	productivity	residual
bialskopodlaski	-20.6	-5.6	3.1	-6.9	-0.5	0.1
białostocko-suwalski	-12.3	-4.8	1.6	-2.7	-5.8	0.3
bielsko-bialski	12.5	5.5	1.8	-0.9	3.0	0.0
bydgoski	0.8	4.5	0.1	-4.4	3.4	-0.2
centralny śląski	29.2	0.2	0.2	5.4	2.3	0.2
chełmsko-zamojski	-19.9	1.6	-0.8	-12.9	-3.0	0.6
ciechanowsko-płocki	-6.9	11.4	-2.1	-14.8	-1.7	0.4
częstochowski	2.8	-6.0	-0.4	-7.4	0.4	0.0
elbląski	-8.0	-0.3	0.1	-10.5	-10.3	1.7
ełcki	-19.0	-4.0	2.0	-15.5	-3.9	1.0
Gdańsk-Gdynia-Sopot	9.0	7.1	1.7	41.2	-0.6	-0.4
gdański	-11.8	2.2	-0.7	-18.4	-2.5	0.7
gorzowski	-1.0	-4.6	0.1	1.6	6.5	0.2
jeleniogórsko-wałbrzyski	-2.2	-0.6	0.0	-15.5	7.6	-1.9
kaliski	-0.5	-9.8	0.1	-13.3	-9.6	2.1
koniński	-0.8	0.0	0.0	-16.8	-4.7	1.3
koszaliński	-11.0	-2.9	0.8	-2.5	-2.9	0.1
krakowsko-tarnowski	-3.9	-3.9	0.4	-19.6	-1.3	0.4
Kraków	17.9	0.8	0.4	49.3	-6.6	-5.2
krośnieńsko-przemyski	-8.5	-1.9	0.4	-11.9	-2.4	0.5
legnicki	13.4	14.8	5.3	-1.2	0.9	0.0
lubelski	-8.3	-2.1	0.5	-2.2	-12.3	0.4
łomżyński	-19.8	-3.8	2.0	-18.8	-6.3	1.9
łódzki	-5.1	-11.1	1.5	-15.7	5.8	-1.5
Łódź	7.5	-2.3	-0.5	24.5	-2.1	-0.8
nowosądecki	-14.8	-0.3	0.1	-13.8	-6.2	1.4
olsztyński	-8.9	-1.7	0.4	2.7	-8.3	-0.4
opolski	0.8	1.2	0.0	-8.8	1.1	-0.2
ostrołęcko-siedlecki	-16.9	-12.5	5.6	-17.1	4.6	-1.3
pilski	-7.1	-7.9	1.5	-10.5	-4.7	0.8
piotrkowsko-skierniewicki	-0.1	-0.5	0.0	-16.8	-8.5	2.3
Poznań	18.6	9.6	4.7	59.1	-3.5	-3.3
poznański	-1.0	-3.4	0.1	-16.9	-5.3	1.4
radomski	-9.5	-10.0	2.5	-18.6	-0.2	0.1
rybnicko-jastrzębski	22.1	0.0	0.0	-10.4	1.8	-0.3
rzeszowsko-tarnobrzeski	1.4	-6.7	-0.3	-8.9	-13.2	1.9
słupski	-7.5	-9.2	1.8	-6.4	9.3	-1.0
szczeciński	-5.7	-1.5	0.2	-10.0	14.9	-2.4
świętokrzyski	-3.8	-1.0	0.1	8.1	-11.9	-1.6
toruńsko-włocławski	-5.8	0.4	-0.1	-8.5	7.7	-1.1
Warszawa	10.2	11.1	3.0	92.0	10.6	15.6
warszawski	-5.1	6.5	-0.9	-15.7	-1.9	0.5
Wrocław	7.5	2.5	0.5	41.0	2.8	1.9
wrocławski	-9.0	-5.0	1.2	-22.0	-1.8	0.6
zielonogórski	-2.7	-6.3	0.5	-2.9	8.1	-0.4

Source: Own calculations based on CSO regional accounts.

Table A.3.
Decomposition of sources of variation in value added per inhabitant from national average - 2004

		industry		services		
	total	employment	productivity	employment	productivity	residual
bialskopodlaski	-15.2	-12.0	5.7	-16.3	-11.8	2.8
białostocko-suwalski	-4.4	-10.0	1.4	-1.9	-11.8	0.3
bielsko-bialski	1.8	9.0	0.5	-9.7	5.2	-0.7
bydgoski	2.9	-3.3	-0.3	-3.0	-3.4	0.1
centralny śląski	13.7	4.2	1.8	6.5	0.1	0.0
chełmsko-zamojski	-11.2	-9.6	3.4	-18.1	-12.2	3.2
ciechanowsko-płocki	-2.8	25.2	-2.2	-9.1	-8.1	1.1
częstochowski	8.6	-3.2	-0.9	-10.4	-3.3	0.5
elbląski	-0.7	-7.1	0.2	-16.8	-7.4	1.8
ełcki	-11.6	-8.0	2.9	-21.5	-4.6	1.5
Gdańsk-Gdynia-Sopot	1.3	5.8	0.2	39.8	2.7	1.6
gdański	-4.4	0.6	-0.1	-24.9	0.8	-0.3
gorzowski	0.4	-4.1	-0.1	-4.7	-4.0	0.3
jeleniogórsko-wałbrzyski	-2.5	-1.7	0.1	-20.8	9.4	-2.9
kaliski	8.9	-9.8	-2.7	-18.5	-5.5	1.5
koniński	3.6	-6.6	-0.7	-21.9	-3.4	1.1
koszaliński	-7.4	-4.5	1.0	-14.2	9.2	-1.9
krakowsko-tarnowski	-4.6	-2.5	0.4	-23.6	-0.6	0.2
Kraków	5.8	5.7	1.0	59.9	-4.9	-4.3
krośnieńsko-przemyski	-7.1	-9.0	2.0	-15.8	-12.8	3.0
legnicki	13.6	15.9	6.8	-4.6	-2.1	0.1
lubelski	-4.9	-5.8	0.9	-0.4	-13.0	0.1
łomżyński	-15.4	2.1	-1.0	-25.6	-5.6	2.1
łódzki	-0.6	-6.2	0.1	-14.2	-8.5	1.8
Łódź	0.8	1.7	0.0	25.7	-0.9	-0.3
nowosądecki	-10.7	-7.2	2.4	-20.9	-8.6	2.6
olsztyński	-2.8	-5.2	0.5	0.8	-8.1	-0.1
opolski	-2.5	4.8	-0.4	-13.5	-3.8	0.7
ostrołęcko-siedlecki	-13.0	-4.1	1.7	-22.2	-5.5	1.8
pilski	4.4	-7.6	-1.1	-18.9	-4.2	1.2
piotrkowsko-skierniewicki	0.6	0.6	0.0	-21.3	-2.9	0.9
Poznań	17.7	6.4	3.6	82.9	1.8	2.1
poznański	11.4	-4.4	-1.6	-17.1	2.8	-0.7
radomski	-9.9	-1.3	0.4	-8.6	-15.0	1.9
rybnicko-jastrzębski	11.0	8.5	2.9	-13.3	-2.7	0.5
rzeszowsko-tarnobrzeski	4.1	-4.8	-0.6	-8.9	-11.3	1.5
słupski	-1.3	-10.2	0.4	-11.4	-5.9	1.0
szczeciński	-6.8	-4.7	1.0	-13.3	9.6	-1.9
świętokrzyski	-5.3	2.9	-0.5	-2.5	-6.7	0.2
toruńsko-włocławski	-1.8	-5.7	0.3	-12.2	-0.9	0.2
Warszawa	4.5	9.5	1.4	122.3	21.0	37.6
warszawski	-4.4	4.7	-0.7	-6.9	1.2	-0.1
Wrocław	-1.2	3.8	-0.1	48.7	-0.9	-0.7
wrocławski	1.1	-1.3	0.0	-25.1	3.1	-1.2
zielonogórski	-0.6	0.3	0.0	-6.5	-4.2	0.4

Source: Own calculations based on CSO regional accounts.

Appendix 2. Methodological remarks on LFS data

In order to estimate the intensity of migration, one can use LFS data basing on the answers to questions about the place of residence one year before the survey. A migrant is then defined as a person who lived in a different district or voivodeship one year before the interview (the survey does not provide for a division at the poviat level). Thus, it is only possible to compare the intensity of migration in Poland with other countries at the NUTS-2 level .

When using LFS data one should take into account the fact that selection of quarterly samples is performed according to the rotation system so the probability of observing migrants in the subsamples might differ. In the Polish survey, one quarter of the sample are households drawn for the first time, one quarter – those which participated in the previous round of the survey (three months earlier), one quarter – those which participated in the preceding year, and one quarter – those which participated 15 months before the examined round (see sample rotation system, Economic Activity of the Population, CSO). This means that for half of the sample the probability that it includes migrants is lower (there might be migration which involved returning to the household after an absence of more than one year or which occured due to the rent or purchase of a flat in the secondary market – whereas people who bought flats in the new place of residence should by no means be included in the sample). Members of households which participate in the survey for the second time (three months after the first round of interviews) are typically characterised by lower probability of migration than those which participate in the survey for the first time. If we ignore the above problem, the estimation of the intensity of migration is biased downwards and the structure of migrant's population is distorted. The LFS sample rotation systems in other countries (Eurostat 2003) indicate that in some countries one third (see France, Spain, Hungary, Greece) or one fifth (see Great Britain, Slovakia, etc.) of the sample is characterised by lower probability of migration related to the purchase of a new flat.

Bearing the above in mind, when comparing migration intensity internationally, one would have to use data from subsamples which participated in the survey for the first time and calibrate weights depending on the rotation system and the number of such subsamples in a given country. For a greater comparability of migration data, the definition of a household, the methodology of determining membership in a household and place of permanent residence as well as of collecting information about people living in collective households should also be standardised.

To some extent the comparability of migration data can be affected by the accuracy of data on new flats, because the sample is drawn based on them.. In Poland, many flats are registered a few years after construction and therefore their occupiers can take part in the survey only a few years after the change of residence.

Moreover, people from villages and small towns who study outside their place of residence and live in halls of residence are not included in the LFS. At the same time - when they graduate and engage in work in the place of study, they are - according to the LFS methodology - classified as people who used to live in a given city for more than one year, i.e. they are not classified as migrants.

Appendix 3. Methodological remarks on the migration model

Model 1: intensity of outmigration from regions

The dependent variable is the intensity of outmigration from particular poviats based on data from the National Population Census on internal migration (Internal population migration, CSO). The intensity of outmigration is calculated as a share of migrants who left a given poviat in the period 1989-2002 in its total population aged 18-59/64 in 2002. At this point, certain reservations concerning the analysis should be made: the above data are subject to an error resulting from the retrospective nature of the migration survey which recorded only the last migration. Moreover, the aggregation of data published by the CSO for the entire period 1989-2002 makes it impossible to establish a panel for poviats and difficult to select explanatory variables (variables describing the situation in poviats are available for 1995-2004). An alternative solution is to use data at the voivodeship level which are disaggregated in time. However, taking into consideration the strong auto-correlation of the basic variables (unemployment rates, wage levels, etc.) for particular poviats in time and the large differences between poviats within voivodeships, it seems better to estimate the model using the data aggregated in time but with little spatial aggregation.

In view of the above (greater "weight" of migration, which occurred at the end of the studied period and strong auto-correlation of explanatory variables) only the data from 2002 were taken into account. The model was reestimated using the data from 1999 – the direction of the interdependence relationship proved the same and only its strength was smaller.

The general formulation of the model is as follows:

$$M_{ri} = f(X_i, Z_i) \qquad \text{where:}$$

M_{Ri} = migration intensity from particular poviats

X_{Ii} = attractiveness of other regions weighted by distance
The "Population Index" was used as a proxy. This index is calculated by multiplying the population in all potential destination poviats by the decay function, derived from the distance from the capitals of these poviats. In spatial mobility models, population of potential destination regions is often used as a measure of their attractiveness.

Z_{Ii} = vector of standardised variables describing the situation in region i

The aim of the study is to measure the impact of the labour market situation on the propensity to migrate to other regions and therefore the set of explanatory variables includes the amount of wages and the ratio of the number of the unemployed to the number of job offers. Attempts were made to include the unemployment rate as a measure of chances for finding employment, however, given the arguments presented by Pissarides and Wadsworth (1989), it seems that the relationship between the number of the unemployed and job offers – irrespective of an imprecision of this statistics –reflects the situation in the region better (what is more this variable was also used to create clusters in Part II). Statistics on the employment structure (shares of people working in agriculture and in services) at the poviat level were also used in the analysis. . In view of the possibility of migration being substituted by work commuting (see Rouvendal 1999), the variables such as number of rooms per capita and road density in poviats were included in the set of control variables (assuming that the greater the supply of flats in a given region, the lesser the tendency of its inhabitants to migrate to the regions of their employment and the greater the tendency to commute – where the latter is further enhanced by greater road density). Explanatory variables (apart from the population index which is "weighted" for all potential destination regions) were standardised working on the assumption that it is not their absolute level but deviations from the average level for all regions that are of key importance.

The model can be formulated as follows

$$M_{ri} = IP * e^{\sum_{i=1}^{k} a_i Z_i}$$

which after the linearization can be estimated by the Ordinary Least Square method:

$$\ln(M_{ri}) = \ln(IP) + \sum_{i=1}^{k} a_i Z_i$$

M_{ri} = log Population Index

Values predicted by the basic model (which included four explanatory variables – logarithm of the population index and standardised values of the supply of flats, number of the unemployed per job offer and wages), are not very well adjusted to the observed values of the intensity of migration in particular poviats However, given the deficiencies of data used in the analysis, this result can be deemed acceptable (road density was also taken into account, which improved the quality of the adjustment but resulted in the collinearity between the population index and road density – in highly populated poviats road density is usually substantially greater. Hence, this

explanatory variable was abandoned). The model explains approx. 44% of dependent variable variation, which should be considered significant given the results of the variation analysis.

The variable which exerts the largest influence on the outflow of migrants from a region is the value of the population index which indicates the availability of other attractive regions. There is a strong negative correlation between the value of the population index and the intensity of migration which seems to be unintuitive but which can be explained by the results concerning the interdependence between population size, distance between regions and the intensity of work commuting. If a less developed region borders with an attractive region which attracts labour resources, workers may prefer commuting to migration given the distance is short. This is so especially when the choice between commuting and migration is additionally motivated by the situation in the real estate market. If flat prices and rents are high in the potential migration destination region, the migration might not be cost-effective even if the wages are relatively high. Because data on flat prices or rents are not available at poviat level, we used statistics on the number of newly built rooms per capita . It turned out that large availability of accommodation encourages people to migrate. It must be specified that this variable may be endogenous, however, we did not find any suitable instrument. What is more, it seems improbable that a small – compared with the total population of a given region – group of work commuters should have a significant influence on the situation in the poviat real estate market. Last but not least, the conclusion that the supply of flats in the region of residence is a factor which affects the choice between commuting and migration and which thus decreases the intensity of outmigration from regions located in the vicinity of attractive regions – is supported by the results of the model of work commuting.

Model 2: Destination choice model

The second model explains the allocation of people who decided to migrated in the period of 1989-2002 between particular regions. The general formulation of the model is as follows:

$$\frac{M_{ij}}{O_i} = \frac{d_{ij}\prod_{i=1}^{k} a_i X_i}{\sum_j d_{ij}} \quad \text{where:}$$

M_{ij} / O_i - share of all migrants originating from the region i who lived in region j in 2002
d_{ij} - function of the distance between regions (reverse)

In view of the fact that the denominator in the equation on the right hand side is the same for each outflow region i , the equation can be transformed to the form which, upon the logarithmisation of both sides, can be estimated using the OLS method:

$$ms_{ij} = k_i d_{ij} \prod_{i=1}^{k} a_i X_i \quad \text{where:} \quad ms_{ij} = \frac{M_{ij}}{O_i}$$

ms_{ij} - all migrants originating from the region i who lived in region j in 2002
k_i - attractiveness of destination regions weighted by distance

In view of the fact that the CSO data on flows below ten people are not available, a tobit model was used and the natural logarithm of 10 was set as the lower limit value. The model is significant, however, it explains only approx. 32 per cent of the variance in the migration flows. All variables apart from the number of the unemployed per job offer are statistically significant and affect the outcome variable in the expected direction.

Appendix 4. World Bank Ease of Doing Business Index

This index allows to rank and rate the economies of 175 countries worldwide for their ease of doing business. Rankings of particular countries on the index are based on the arithmetic mean derived from their rankings in each of the ten examined areas. Country rankings are expressed as percentiles within which the result for a given country falls.

The Ease of Doing Business Index is based on the following subindices:

- **starting a business** – records all procedures that are required for an entrepreneur to start up and operate a business; these include obtaining all necessary licences and permits and completing any required notifications, verifications or inscriptions for the company and employees with relevant authorities, costs of complying with the above procedures and minimum capital requirements;

- **dealing with licenses** – records all procedures required for a business in the construction industry to build a warehouse of a given size (including time necessary to obtain all licences and permits, to receive all required inspections and to complete and submit the relevant documents); the index also records procedures for obtaining utility connections (water, electricity, etc.) and time necessary to obtain a permit for transferring or selling the said warehouse;

- **employing workers** – records activities that employers are obliged to perform when hiring or firing employees, as well as costs relating to dismissals and rigidity of working hours;

- **registering property** – records all procedures necessary when a business purchases land and buildings to transfer the property title from the seller to the buyer so that the buyer could use the property; all procedures are included irrespective of whether it is the responsibility of the seller or the buyer or must be completed by a third party on their behalf;

- **getting credit** – constructs measures of the legal rights of borrowers and lenders, describes how well bankruptcy and collateral laws facilitate landing and indicates how to access credit information (loan terms and conditions and information about borrowers) available through public and private credit registers;

- **protecting investors** – measures the strength of minority shareholder protection against directors' misuse of corporate assets for personal gain; the index distinguishes three dimensions of investor protection: transparency of transactions, directors' liability for self-dealing (profits derived from confidential information) and shareholders' ability to sue officers and directors for misconduct;

- **paying taxes** – records taxes which a medium-size company must pay or withhold in a given tax year, as well as to administrative burden in paying taxes with respect to all components of the tax burden (corporate income tax, social insurance contributions and other labour taxes, property taxes, property transfer – including sale – taxes, dividend tax, capital gains tax, financial transactions tax, road taxes and waste collection taxes);

- **trading across borders** – compiles all procedural requirements for exporting and importing goods – from the contractual agreement between the seller and the buyer to the delivery of goods, along with time and cost necessary for completion;

- **enforcing contracts** – measures the efficient of the judicial system in resolving commercial disputes;

- **closing a business** – studies the time, cost and outcomes of bankruptcy proceedings involving domestic entities.

Appendix 5. Paying Taxes

For each country included in the Ease of Doing Business Index, the Paying Taxes Index was estimated using a case study with a set of financial statements and assumptions about transactions made over the year. Experts in each country calculate taxes to be paid in their jurisdictions. Moreover, information was gathered about the frequency of tax return filings, audits and other costs of compliance with tax obligations.

The index was computed for a business with certain set characteristics, identical for all countries. The business:

- is a limited liability, taxable company,
- started operation on 1 January 2004; at the same time the company purchased all the assets shown in its balance sheet and hired all its workers,
- operates in the country's most populous city,
- is 100 per cent domestically owned and has five owners, all of whom are natural persons,
- has a start-up capital of 102 times income per capita at the end of 2004,
- performs general industrial and commercial activities; specifically, it produces ceramic flowerpots and sells them at retail,
- does not participate in foreign trade and does not handle products subject to a special tax regime (e.g. alcohol, tobacco),
- owns two plots of land, one building, machinery, office equipment, computers and one truck and leases another truck,
- does not qualify for investment incentives or any special benefits apart from those related to the age or size of the company,
- had sixty employees – four managers (including one owner), eight assistants and forty-eight workers – all nationals,
- has a turnover of 1,050 times income per capita,
- makes loss in the first year of operation,
- has the same gross margin (pre-tax) across all economies,
- distributes 50 per cent of its profits as dividends to the owners at the end of the second year,
- sells one of its plots of land at a profit during the second year.

Moreover, relevant assumptions were also made about taxes. All the taxes were paid in the second year of operation, i.e. in 2005. All the taxes paid over the year were taken into consideration, however, taxes with the same name and taxes collected by the same agency (though charged at different rates) are counted as the same tax. The frequency of payment for taxes includes advance payments or withholding as well as regular payments or withholding.

The index consists of three elements:

- tax payments (total number of taxes paid, method of payment, frequency of payment and number of agencies involved);
- time – recorded in hours per year – to collect all information necessary to compute the tax payable, to file a tax return and to pay the amount due for three major types of taxes: corporate income tax, value added tax or sales tax and labour taxes, including payroll taxes and social insurance contributions;
- total tax rate which measures the amount of taxes payable by the business in the second year of operation, expressed as a share of commercial profits (sales minus cost of goods sold, minus gross salaries, minus deductible provisions, etc. plus capital gains from the sale of property).

Appendix 6. Employing workers

This index measures the regulation of employment, specifically as it affects the hiring and firing of workers and the rigidity of working hours. To make the data comparable across countries, several assumptions about the worker and the business are used. The information on employing workers is based on a detailed survey of applicable employment regulations that is completed by local law firms. Laws and regulations as well as secondary sources are reviewed to ensure accuracy. Conflicting answers are further checked against two additional employment regulation sources.

The index applies to a worker who:

- is a non-executive, full-time male employee who has worked in the same company for 20 years,
- earns a salary plus benefits equal to the country's average wage during the entire period of his employment,
- is a lawful citizen with a wife and 2 children,
- resides in the country's most populous city,
- is not a member of a labour union, unless membership is mandatory.

The business which employs the worker:

- is a limited liability company,
- operates in the country's most populous city,
- is 100 per cent domestically owned,
- operates in the manufacturing sector,
- has 201 employees,
- abides by every law and regulation but does not grant workers more benefits than what is legally mandated.

The Employing Workers Index consists of three indices:

- the rigidity of employment index which is the average of three subindices: (i) difficulty of hiring index (i.e. whether term contracts can be used only for temporary tasks, what is the maximum cumulative duration of term contracts, what is the ratio of the minimum wage for a trainee or first-time employee to the average value added per worker), (ii) rigidity of hours index (i.e. whether night and weekend work is restricted, whether the workweek can consist of 5.5 days or 50 hours for two months a year, what is the length of paid annual vacation), and (iii) difficulty of firing index (i.e. whether redundancy is disallowed as a basis for terminating employment relationship, whether the employer needs to notify a third party to terminate one redundant worker or a group of at least 25 redundant workers, whether the employer needs approval from a third party to terminate one redundant worker or a group of at least 25 workers, whether the law requires the employer to consider reassignment or retraining options before redundancy termination, and whether priority rules apply for redundancies and reemployment);
- he non-wage labour cost index which measures all social insurance payments and payroll taxes associated with hiring an employee in fiscal year 2005, which are expressed as a percentage of the worker's salary;
- the firing cost index measures the cost of advance notice, severance payments and penalties due when terminating a redundant worker, expressed in weekly wages.

Appendix 7. Rule of Law and Government Effectiveness

The Rule of Law Index and the Government Effectiveness Index are two of six indices developed to measure governance in particular countries (Governance Indicators) and published by the World Bank. These indices are computed based on surveys and polls carried out by different international institutions such as the World Bank, Gallup International, Heritage Foundation, Reporters Without Borders (for a detailed description of data sources see Kaufmann et al. (2006)). Data from particular sources are standardised and aggregated into indices using an unobserved components model (see Kaufman et al. (2004)). Higher values indicate greater governance.

The Rule of Law Index measures the extent of compliance with the rules of society and the inclination to abide by them. In particular, this index reflects the quality of contract enforcement, the police, and the courts, as well as the likelihood of crime and violence.

The Government Effectiveness Index studies the quality of public service provision, the competence of civil servants and the independence of the civil service from political pressures. Moreover, this index measures the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies (see Kaufmann et al. (2006)).

Appendix 8. Heritage Foundation Regulations Index

The Regulations Index constitutes one of measures used to compute the Index of Economic Freedom published by the Heritage Foundation. In accordance with the methodology applicable in 2003, the index could take discrete values within a 1-5 performance bracket. The lowest results mean that the regulatory burden in a given country is not significant, regulations are transparent and they do not discriminate any particular type of business and that the scale of corruption is small. On the other hand, the highest index value characterised countries where government regulations hindered the ability to start new businesses, where corruption had grown to considerable proportions, regulations were non-transparent and selective.

Index values were computed for particular countries based on a set of quantifying variables, such as the extent of licence requirements when starting a business and the ease of obtaining licences, the scale of corruption, the extent of labour regulations as well as other regulations which are a burden on businesses. Data was derived, among others, from reports published by the Economist Intelligence Unit and from publications prepared by national governments. Relevant descriptions were ascribed to every index value. The condition for classifying in a given category was the fact of having most features identified with a given index level.

Appendix 9. Transparency International Corruption Perceptions Index

The Corruption Perceptions Index is based on survey and questionnaire results as well as interviews carried out by international institutions. The 2003 value of the index was computed from 17 variables derived from 13 sources (including World Bank, EBRD, World Economic Forum). All sources used to compute the index referred to the same definition of corruption. Questions related to subjective perceptions of the frequency of corruption in the closest environment, in a given sector or on the national level. All sources provided a ranking of countries. The values were compiled and standardised with the use of statistical techniques so that results would remain between 0 and 10. Higher index values are typical for lower extent of corruption in a given country.

Appendix 10. Sample selection model

In the situation where the selection of observations to the sample is non-random, there may arise a problem of quality of parameter estimates. The estimated values may provide a distorted image of the direction and strength of the relationship between the explained variable and the explanatory variables. For instance, estimation of the traditional regression model which quantifies the impact of secondary school performance on performance at the university level, may lead to the conclusion that there is no significant relationship between these variables. However, **in population**, an evident relationship can be observed between educational achievements at particular steps of the education ladder. The lack of such correlation in the model results from the fact that the probability of going to university, which is practically tantamount with being included in the examined sample (not all people who enter universities actually graduate), depends on the secondary school performance. Good performance in secondary school increases the chances of being admitted to university and thus of completing university studies. Similarly, attempts at determining the impact of particular factors on the probability of shadow employment in the population may be distorted due to different inclinatons to engage in hired work by members of different social and economic groups.

Non-random sample selection makes the OLS estimator inconsitent. Let us assume that we are estimating parameters of the following equation:

$$\mathbf{y}_i = \mathbf{x}'_i \beta + \varepsilon_i$$

where sample selection is determined by the following equation:

$$\mathbf{z}_i = \mathbf{w}'_i \gamma + \upsilon_i$$

The value of the explanatory variable is observed when $z_i > 0$. The correlation coefficient between disturbance terms is ρ. The regression function in the observed sample takes the following form:

$$E[\mathbf{y}_i \mid \mathbf{z}_i > 0] = E[\mathbf{y}_i \mid \upsilon_i > -\mathbf{w}'_i \gamma] = \mathbf{x}_i' \beta + \rho\sigma_\varepsilon \lambda_i(\mathbf{w}_i, \gamma, \sigma_\upsilon)$$

where λ_i means a function of explanatory variables and selection equation parameters, and σ_υ means a standard deviation of the random disturbance ε_i. Clearly the above equation differs from the regression equation in the population which is as follows:

$$E[\mathbf{y}_i] = \mathbf{x}_i' \beta$$

Hence, the estimation of the above function on the observed sample gives rise to the specification error. The β estimator is not consistent in this case (see Greene (2000)).

The most commonly used estimation method when dealing with non-random sample selection is Heckman's two-step method (see Heckman (1979)). In the first step, the probit model is estimated which describes sample selection and determines λ_i values for particular observations. In the second step, it is the parameters of the main equation that are estimated and this equation also takes the form of probit regression.

Appendix 11. Methodological remarks concerning the analysis of internal migration and work commuting

Internal migration

Model of individual determinants of internal migration was estimated using Polish LFS data. A migrant was defined as a person who had lived in a different district or voivodeship one year before the interview (LFS does not provide for a division at the poviat level).

When using the LFS data to analyze migration, it should be borne in mind that due to the sample rotation system the probability of observing migrants in the subsamples might differ. (See Appendix 2). Members of households which participate in the survey for the second time (three months after the first round of interviews) are typically characterised by lower probability of migration than those which participate in the survey for the first time. If we ignore the above problem, the estimation of the intensity of migration is biased downwards and the structure of migrant's population is distorted.

In order to eliminate the above-mentioned problem, in the logit model of migrations we used LFS data only from the first quarters of 2001-2005. In order to obtain a balanced sample, we included all the observations on migrants and we drew a subsample of people who did not changed their place of residence. The model was estimated using Maximum Likelihood Method on the subsample of 4,284 observations.

Work commuting

Model of individual determinants of work commuting, was estimated using the data from the European Social Survey (ESS). The total ESS sample consisted of 45,000 observations, however, we only included people in the productive age, commuting to work and living in the EEA – hence, the analysis was based on 11,044 observations. The model was estimated using Maximum Likelihood Method.

Abbreviations used in the report

APW – average production worker income

CEE – Central and Eastern Europe

CPI – consumer price index

CSO – Central Statistical Office

DEA MLSP – Department of Economic Analyses and Forecasts of the Ministry of Labour and Social Policy

EBRD – European Bank for Reconstruction and Development

EEA – European Economic Area

EMU – Economic and Monetary Union

EPL – employment protection legislation

ERM II – Exchange Rate Mechanism II

ESA – European System of Accounts

ESS – European Social Survey

EURES – European Employment Service

FDI – foreign direct investment

GGDC – Groningen Growth and Development Centre

HCPI – harmonised consumer price index

IBS – Institute for Structural Research

IDA – Industrial Development Agency

ILO – International Labour Organisation

IMF – International Monetary Fund

LFS – Labour Force Survey

MLSP – Ministry of Labour and Social Policy

NBP – National Bank of Poland

NMS – New Member States; admitted to the EU on 1 May 2004

NOE – non-observed economy

NPC – National Population Census

NUTS – Nomenclature of Units for Territorial Statistics

OECD – Organisation for Economic Cooperation and Development

PCA – Polish Classification of Activities

PFS – public finance sector

PPS – purchasing power standard)

RDB – Regional Data Bank

SAE – small area estimation

SEZ – Special Economic Zone

SII – Social Insurance Institution

TFP – total factor productivity

TUS – Time Use Survey

WRS – Worker Registration Scheme