Distr.
GENERAL

Working paper 11
27 February 2008

ENGLISH ONLY

**UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)**

Joint UNECE/Eurostat Work Session on Migration Statistics
Geneva, Switzerland, 3-5 March 2008

Item 4 of the provisional agenda

CHALLENGES IN DEFINING AND MEASURING DIFFICULT-TO-COUNT MIGRANTS

METHODOLOGIES FOR THE ESTIMATION OF STOCKS OF IRREGULAR MIGRANTS*

I. INTRODUCTION

1. Notable fields of public policy where knowledge is inherently limited include homelessness, drug abuse, tax evasion, corruption, cigarette smuggling, drunk driving, under-age prostitution, stigmatizing diseases, mental illness, the "informal economy" as well as illegal migration.[1] All these areas involve so-called "hidden populations" or "hidden activities" that are either difficult to observe or, once observed, are difficult to identify as belonging to that population or as performing such activities. For example, persons with certain characteristics (such as being illiterate, paranoid or having HIV/AIDS) or experiences (such as having become homeless or a victim of rape) may wish not to be identified out of fear or shame and often seek to hide it (cf. Chelimsky 1991, p. 685f). Moreover many, though by far not all of these hidden phenomena, involve illicit activities that make an objective counting and description of them even more difficult. However, for various reasons, knowing more about the size, characteristics

* This paper has been prepared by Michael Jandl, on behalf of Medstat II programme, at the invitation of the secretariat.

[1] Note that the terms "illegal migration" and "irregular migration" are often used interchangeably in most policy contexts. Strictly speaking, however, "illegal migration" refers to the illegal crossing of borders only, while "irregular migration" covers a wider area of irregularities in the status of migrants (e.g. illegal residence after the expiry of visas (visa overstayers) or the illegal employment of foreigners in an otherwise "regular" situation). Cf. Jandl and Kraler (2006, p.337f).

and social behaviour of such "hidden populations" is of considerable interest to policy-makers as well as social scientists, who are attempting to tackle the challenge of researching such "hard-to-reach" or "difficult" populations.

2. Generating knowledge on the size and composition of irregular migrants is not much different from generating estimates on other "hidden populations" mentioned above and thus it is also plagued by many of the same methodical problems. A full count of the population of "undocumented" migrants is, by definition, impossible. It is also not possible to draw a representative sample from the total population, as the structure of the underlying total is unknown. In fact, without intrusive (police) methods it may even be quite difficult to identify whether a *particular* person is indeed an irregular migrant, even when observed and questioned. Like people involved in other illicit activities, irregular migrants have incentives to deliberately hide from public authorities.

3. Thus, the number of irregular migrants that are documented (in one way or another) at any one time is inevitably only a subset of the total population of irregular migrants while the "true number" of irregular migrants can never be known with certainty or any degree of precision. However, public policy needs to be guided by evidence, often in the form of facts and figures. This leads to calls for estimating the "*dark figure*" of official statistics, a term that designates that part of the irregular migrant population that is not documented in the data but is likely to constitute the major part of it. As we will see in the following sections, there are various methodologies for estimating irregular migrant populations.

II. A CLASSIFICATION OF METHODS FOR THE ESTIMATION OF IRREGULAR MIGRATION

4. In analogy to data on legal migration, the fundamental distinction between all estimates on irregular migration is that they refer to one of two distinct statistical concepts: stocks (e.g. of undocumented/illegal residents or irregular migrant workers at a point in time) or flows (e.g. of illegal entrants or migrants "overstaying" over a certain period). While theoretically the two concepts are linked, in practice the quantitative consistency of the two variables has proven to be elusive. Moreover, given the highly volatile nature of migration flows, the scarcity of reliable indicators on illegal migration flows, and the dearth of appropriate methods for estimating such flows, most efforts have concentrated on estimating stocks of undocumented migrants rather than flows. Accordingly, the following discussion focuses on available methods for estimating the size of stocks of undocumented residents with only few references to their interlinkages with flows.

5. Methods for the estimation of stocks of illegal residents can be divided into direct and indirect approaches.[2] Direct approaches are based on data that "capture" the subject of research

[2] It should be noted that the classification proposed here has some similarities but also significant differences to the typology proposed by Delaunay and Tapinos (1998a, p.36ff) and taken up in the excellent overview provided by Pinkterton et al. (2004, p.33ff) due to differences in the criteria used for classification. For example, in the classification suggested here, the Delphi-method is not classified as a direct but an indirect method (see text).

(i.e. illegally resident foreigners) directly, while indirect approaches do not rely on such data. Data sets that contain (a subset of) the target population directly as (identified) illegal residents are immigration enforcement data (e.g. apprehended illegal residents), administrative records (e.g. data on regularization of unauthorized residents) and survey data (e.g. illegal residents identified through snowball sampling techniques). On the other hand, residual methods based on the difference between, for example, the total population represented in census figures and some estimate of the legally resident population are classified not as direct but indirect approaches, because at no point do they identify illegal residents directly in their counts.

6. Direct estimation approaches can be further classified into multiplier methods, advanced statistical methods, methods of self-identification and snowball sampling (referral by others) methods. For each of these categories further sub-categories (for example, among the multiplier methods, there are simple multiplier models, capture-recapture models and models using a comparison of administrative registers) can be identified. Among the indirect approaches, we can identify residual methods, demographic methods, subjective estimations/indicators methods, econometric methods on the size and structure of shadow economies, comparisons of immigration and emigration statistics, flow-stock methods and methods based on indirect inferences as the most import estimation techniques. There are also several combined approaches that use a combination of data sources or estimation techniques. Table 1 and Table 2 provide an overview of this classification scheme for stocks and flows, respectively.

Moreover, over the past decade, there have also been new methods developed and described in the literature that have not been available before.

Table 1: A classification of methods for the estimation of stocks of irregular migrants				
Approach	**Data sources**	**Method**	**Estimation technique**	
Direct approaches	Based on immigration enforcement data	Multiplier methods	Simple Multiplier	1.
			Capture-recapture	2.
			Repeated capture	3.
			Matching of registers	4.
		Statistical methods	Random effect mixed modelling approach	5.
	Based on administrative statistics	Methods of self-identification	Evidence based on regularisation data	6.
			Using data on status adjustments over time	7.
	Based on surveys	Snowball sampling	Direct survey methods	8.
			Single stage link-trace sampling	9.
Indirect approaches	Based on census/registers	Residual methods	Differences census results – legal immigration data	10.
			Simple comparison of various registers	11.
		Demographic methods	Use of birth/death rates	12.
	Based on surveys of "key informants"	Subjective Estimations/ Indicators Methods	Expert surveys	13.
			Delphi surveys	14.
	Based on non-demographic data	Econometric methods on shadow economy	Inference from estimates on illegal work	15.
	Based on census/registers/ demographic data	Expected population methods	Comparison of census/emigration data and immigration statistics	16.
	Based on administrative statistics	Flow-stock methods	Calculating the stock through flow figures	17.
	Based on complementary data sources	Indirect inferences	Registered school children, household surveys, remittance data, etc.	18.
Combined approaches	Based on small scale surveys	Window/Postal code method	Small scale study / use of regression analysis	19.
	Based on expert opinions	Localized Delphi	Delphi method / use of regression analysis	20.
	Adjustment to surveys/ census data	Non-threatening survey design	Randomized response (3 cards method) / residual method	21.

Table 2: A classification of methods for the estimation of flows of irregular migrants				
Approach	**Data sources**	**Method**	**Model**	
Direct approaches	Based on border apprehension data	Multiplier methods	Simple multiplier	22.
Indirect approaches	Based on stock estimates	Differential methods	Net differences in stocks	23.
	Based on entry-exit statistics	Residual method	Double entry card system	24.

III. DIRECT APPROACHES

A. MULTIPLIER METHODS

1. Simple Multiplier Methods

7. Among the methods used for the estimation of stocks of undocumented migrants, a significant share is based on the "multiplier principle". This method starts from the proposition that the size of the unknown total can be directly calculated from the size of a known subtotal by use of an appropriately estimated multiplier (for example, that the stock of irregular migrants in a country at a given time can be derived by use of a multiplier on the number of detected irregular migrants). Once this proposition is accepted (and the size of the subtotal has been established with an acceptable degree of accuracy), the problem is redefined as finding the "right" multiplier.[3]

8. The use of multipliers to derive the size of a hidden population from the size of a known subtotal of that population is probably the most common method of estimating the "dark field" of an unknown population in this as in other fields. For example, the United Nations Office on Drugs and Crime uses various multipliers to derive the size of problem drug users in a country: "If a survey among heroin addicts reveals, for instance, that one quarter of them was in treatment in the last year, the multiplication of the registered treatment population with a multiplier of four provides an estimate of the likely total number of problem heroin users in a country" (UNODC 2007, p. 266). An alternative estimate that can be compared to the first one is derived by applying a multiplier on arrest data: "...if a survey among heroin addicts reveals that one out of five addicts was arrested in the previous year, a multiplication of the persons arrested for heroin possession by the multiplier (five) provides another estimate for the number of heroin users" (ibid).

9. A good example of using a sample of the total population for estimating a multiplier is provided by Burgers (1995, 1996 cited in Pinkerton 2004, p. 14f) who uses a record and survey-based method of determining the number of illegal foreigners in Rotterdam at a particular moment in time. First, Burgers uses the number of apprehended foreigners over a six year period to gain a measure of how many criminal illegal foreigners were captured. Then he uses in-depths interviews with a (non-random) sample of illegal migrants (145 interviews) to determine the share of migrants involved in criminal activities. This proportion was then applied to the apprehension number derived before to produce an estimate of the total illegally resident population in Rotterdam. Finally, the share of the estimated population in the total population (1.8 %) was used to extend the estimate, first to the four largest cities in the Netherlands (Amsterdam, Rotterdam, The Hague and Utrecht) and second to the whole country (In extending the estimate from the four largest cities to the whole country, a lower share of illegal residents in rural areas was assumed; this spatial concentration of irregular migrants in large cities and in

[3] Vogel, D. (2002): Ausländer ohne Aufenthaltsstatus in Deutschland. Methoden zur Schätzung ihrer Zahl. In: Minthe, Eric (ed.): Illegale Migration und Schleusungskriminalität. Wiesbaden, Eigenverlag der Kriminologischen Zentralstelle e.V., 2002, p. 70

specific rural area is an important aspect of the demographic composition of illegality in European countries (cf. Leerkes et al 2007).

2. Capture-Recapture Methods

10. Capture-Recapture methods have a long tradition in population biology and were originally developed to estimate animal populations in the wild (Petersen 1896). Today, capture-recapture methods are widely used in epidemiology and in the estimation of hidden populations.

11. In their simplest form, capture-recapture methods are a sort of multiplier method where the multiplier is developed through repeated sampling of the same population. To illustrate, consider the following application of the principle to the estimation of the stock of fish in a pond. First, capture 1,000 fish, mark them, and let them free again. Then, capture another 1,000 fish and examine them. If 100 of them are marked (recaptured), you can deduce that the 1,000 marked fish statistically make up 10% of the total, so there are – presumably – 10,000 fish in the pond (cf. Jandl 2004).

12. There are few studies that apply this classic capture-recapture technique in a migration context.[4] However several studies apply the principles of this method in an advanced form to the estimation of the stock of illegal residents (see 4.1.3 below).

3. Repeated Capture Methods

13. Based on the above principle, a mathematically more demanding variant of the capture-recapture method has been applied for the first time with regard to illegal residents in the Netherlands. Van der Leun, Engbersen and van der Heijden (1998) use the so-called "repeated capture method" for estimating the size of the illegally resident population in the four largest Dutch cities.[5] The "repeated capture method" is based on one single data set (7,000 files related to all apprehensions of illegal immigrants in 1995 in the four largest Dutch cities) in which illegally resident foreigners may appear more than once. By analysing the files, it is determined who is captured once, twice, three times and so on.[6] Using the counts of persons captured and re-

[4] However, the method has recently become used in the estimation of the number of victims of human trafficking (both internal and international). The widely cited ILO estimate of global human trafficking victims is based on a) an estimation of reported cases worldwide by way of a capture-recapture technique that seeks to establish the number of *reported* cases through comparing two samples of reported cases in order to estimate all reported cases not captured in the samples; and b) a further extrapolation of this estimated number by a factor of 10 to arrive at the ILO global minimum estimate of forced labour and, as a subgroup, the ILO global minimum estimate of human trafficking. See Belser, P., de Cock, M. and Ferhard M. (2005), ILO Minimum Estimate of Forced Labour in the World, ILO, Geneva, April 2005

[5] The study was part of a wider study known as the "Unknown City Project". Their estimation results (a minimum number of 40,000 illegal residents in Amsterdam, Rotterdam, The Hague and Utrecht together) are surprisingly close to the estimate provided by Burger 1996.

[6] An important precondition for applying the method is the ability to observe the same persons at least twice. As only few people will be captured twice and even fewer more than twice, the results of the estimation depend crucially on a correct identification of persons "recaptured" and a good matching of records, which is, however, made more reliable through the use of fingerprinting.

captured 1, 2, 3 and so on times, it is postulated that the number of appearances follows a probabilistic distribution, the Poisson distribution. On the basis of the available data, the crucial parameter determining the Poisson distribution (usually denoted by the Greek letter Lambda) can be estimated, which is then used to calculate the probability that an individual is never (zero times) caught by the police. Adding up this estimated number with the number of illegal residents actually "captured" various times in the police files produces an estimate of the total illegally resident population (cf. van der Heijden, Bustami et al 2003, van der Heijden, Cruijff et al 2003).[7]

14. The repeated capture method, as originally applied in the Netherlands depends on a number of crucial assumptions that can only be briefly mentioned here. First, the chance of getting caught over the period of study (one year in this case) remains constant and this probability does not change after a previous capture.[8] Second, the total population under consideration remains constant over the period (no inflows or outflows). Obviously, this is not the case at least for those apprehended illegal residents who are subsequently removed from the country, therefore a separate estimation is performed for the group of effectively expelled persons (van der Heijden, Bustami et al 2003). Finally, when applying a uniform probability of being caught for all persons, the underlying assumption is that the total population is homogeneous with respect to their risk of being caught. However, van der Heijden (2006) demonstrates that this assumption can be relaxed when differences in the risk of being caught are included in the model by explicitly estimating different Poisson parameters for different groups of illegal residents.

4. Matching of Registers

15. At first sight, this estimation technique may not appear to be a variant of capture-recapture methods but on closer inspection it becomes clear that it belongs to that group. It uses the fact that illegal residents sometimes appear in a certain registry (e.g. police apprehension files, aliens registry) and can be identified and re-identified when appearing in another registry as well. If the probability of appearing in one register is independent of that appearing in the other one, the identification of the same individual in both registers constitutes, metaphorically speaking, a "recapture" of the same individual and the total can be derived from the sizes and capture rates of the two samples.

16. Like the other capture-recapture methods, the application of this estimation technique is based on a number of crucial and partly problematic assumptions. First, the method assumes a closed population for the period the estimates relate to and no linking errors (i.e. a person can be correctly identified as appearing in both registers or only one register). Two further assumptions – that of homogeneity (all persons must have the same probability to be "captured") and independency (the probability to be captured in one of the two registers is not influenced by the

[7] For a short description of the method and results, see Pinkerton et al. 2004 and Sikkel et al. 2006

[8] This also means that changes in policy or law enforcement efforts that could affect this probability are only insignificant over the period.

fact of being captured in the other registry) can be corrected for by advanced statistical methods with the use of additional co-variates and registries (cf. Sikkel et al 2006).

B. STATISTICAL/PROBABILISTIC METHODS

17. These methods use sophisticated statistical inference methods on certain (usually large) datasets on irregular migration phenomena (e.g. police apprehension data) to draw statistically relevant conclusions on the size of the illegally resident population.

1. A Random Effects Mixed Modelling Approach

18. This approach tries to estimate the size of the illegally resident population through statistical modeling from data on registered legal residents (N) and data on apprehended legal (n) and illegal (m) residents to estimate the number of non-registered illegal residents (M). Thus, while the proportion of apprehended foreigners is known among the legal residents (i.e. n/N), it is unknown for the illegal residents (i.e. m/M is not known). The model realistically assumes that there are a number of reasons why these rates must differ and without knowing the difference one cannot estimate M directly. However, the model introduces a random effect that is able to accommodate the perceived heterogeneity in the relative apprehension rates for individuals and makes the estimation of M possible.

19. In applying the model to data on registered and apprehended foreigners in Norway, Zhang (2008) finds that the regression of the model to the data results in a good fit and thus allows for an estimation of the illegally resident population within reasonable confidence intervals.

C. METHODS OF SELF-IDENTIFICATION

20. Methods that rely on the self-identification of members of “hidden populations” are often mistaken for counts of the total. Regarding irregular migration, in practice the best known example of an “estimation method” that relies on self-identification of undocumented migrants is the count obtained from large-scale regularizations. These data are often interpreted as an approximate size of the illegally resident (or illegally working) population with the implicit assumption that all members of the hidden population would be able and willing to take advantage of an uncertain chance to obtain a regular status in return for revealing their identity and (work or residence) status to the authorities. As in most cases no explicit estimation of the underlying total is made, it is better to use the term “evidence” rather than “method”.

1. Evidence based on Regularization Data

21. Large scale regularizations (or amnesties) are often carried out to deal with the perceived problems of dealing with a large number of irregular migrants or migrant workers. Almost as a by-product they also provide information on the illegally resident population (cf. Pinkerton 2004, p.39f).

22. Depending on the terms of the regularization on offer, the number of persons applying may be a good indicator of the total number of persons in the relevant category (illegal residents,

irregular foreign workers, etc.) but there are also good reasons to expect significant deviations from the total number of persons in that category. First, not all eligible persons will apply for a regularization of their status. In most cases, the number of regularizations granted is significantly lower than the number of applications and not everyone will be willing to take the chance. Second, the number of applications is sometimes inflated when the same persons apply more than once to increase their chances. Thus the raw application data have to be cleaned of double-entries.[9] Third, when the regular status (residence or work) is offered only for a relatively short period of time (e.g. one year), once regularized migrants may quickly fall back into irregularity (thus showing up repeatedly in successive amnesty programs.[10]

23. Finally, instead of pertaining only to a defined group of irregular foreigners in the country the regularization is carried out, an unknown number of (irregular) foreigners from other (usually neighbouring) countries may also file an application.[11]

D. METHODS USING SNOWBALL SAMPLING

1. Direct Survey Models

24. Snowball sampling as a means to estimate the size of a hidden population is fraught with many difficulties and potential sample biases. Natale (1998) reports on various examples where the method has been tested in Italy and distinguished between different types of snowball sampling. The method proposed by himself starts with the extraction of an initial sample from a list of registered foreigners and asks each interviewee to indicate another foreigner (who is either in the initial sample or not) to obtain a second sample and so forth. Taking all the interviewed persons *not* in the initial sample together one can obtain an estimator of the probability of not belonging to the initial list and from there an estimate of the non-registered foreigner population. Natale admits that the method has several defects and introduces two variants of the method. The first is the so-called habitation snowball method, which takes designated (fixed) houses for immigrants as sampling places for foreigners, collecting also information on co-habitants not present. The second method proposed is the so-called hospitality centre or meeting place snowball, which makes contact with migrants (regular and irregular) at popular centres of hospitality (churches, etc.) and meeting places of migrants. Questioning the migrants for their usual meeting places/centres frequented by them a matrix of all such centres visited by certain interviewees can be drawn up and analysed for statistical frequencies to obtain an estimated total of the whole group (see Natale 1998, p.10).

---

[9] Pinkerton et al 2004, p. 40, cite the example of the French amnesty 1997, where the number of applications were thus reduced from 180,000 applications to 144,000 applicants.

[10] Depending on the terms of the regularization it is likely that the extent of this phenomenon varies. For example, in the case of Italy relatively few are thought to fall back into irregularity while in the case of Greece the majority of beneficiaries of recent regularization programs are thought to have fallen back into an irregular status. Cf. OECD (2004), p.70

[11] For example, in Italy it has been observed in past regularizations that a number of irregular migrants normally residing in France tried to obtain a regular status by applying for the regularization in Italy moving to Italy just to file their application. Sciortino, G. (2003): Regularization of Foreigners in Italy, unpublished manuscript, Communication to Michael Jandl on 3.4.2003, p.2

2. Single Stage Link-Trace Sampling

25. The snowball method described above is a non-random sampling method that relies on a number of initial respondents for locating additional respondents of a hard-to-reach population (such as drug users or illegal residents). This method is conventionally used for identifying a larger number of respondents (of hidden populations) for qualitative interviews.[12] There are also attempts to apply the method to quantify the size of a hidden population through statistical inferences. Zhang (2007) provides an overview of the data requirements and the calculations for producing a statistical estimate of the so-called *single-stage link-tracing sampling*, the simplest form of quantitative snowball sampling. In this model, initial respondents are asked to nominate other members of the target population (one round only). Nominees do not have to be known by name but instead the requirement is only to identify whether they are in the sample of initial respondents or not. The higher the (relative) number of persons nominated that link back to the initial sample of respondents, the smaller the predicted total of the target population. A formula can be used to calculate the estimated total.

26. While the data requirements may not seem to be very high, in reality they are likely to work only for relatively small target populations. Zhang (2007) provides the example of the study of Frank and Snijders (1994) who used the method to estimate the number of cocaine users in Rotterdam, a relatively small hidden population.[13] Given these constraints, the method has so far not been used for sizing relatively large irregular migrant populations. However, the method may be fruitfully applied in combination with other estimation techniques, e.g. methods suitable for extrapolating from smaller to larger areas.

IV. INDIRECT APPROACHES

27. These methods (sometimes also called “derivative methods”) use the postulated correlation of the unknown variable (e.g. the size of the illegally resident population) to a variable that can be measured or estimated by other means (e.g. the size of the informal economy) for making inferences on the size of the unknown variable. The inference from the directly measured variable to the indirectly measured quantity necessitates further assumptions and (often) estimated relationships. There are great differences between the methods subsumed under “indirect methods”. What they have in common is that at no point do they use counts of irregular migrants (such as apprehensions etc.) directly for their estimations.

[12] It may often also be the only method available for identifying a sizeable sample of a hidden population for the purpose of estimating a multiplier, as in the study of Burgers (1995, 1996 cited in Pinkerton 2004, p. 14f).

[13] From an initial sample of 34 persons, the number of nominations was 311 of which 15 pointed back to the initial sample. This yielded an estimated total of 685 persons in the target population.

A. Residual Estimation Methods

1. Comparison of Census Results with Legal Immigration Figures

28. This method is based on the proposition that a total count of the immigrant population in a country would capture both legal and illegally resident immigrants at the same time, while not allowing a distinction between them. Subtracting the total number of legally resident immigrants from this total count then gives the estimated number of illegally resident immigrants.

29. This method is most commonly applied in the United States, where the Census (carried out every ten years) is thought to give a good count of the total foreign born population, including the illegally resident population (cf. Pinkerton 2004, p.20f). Undercounting in the census in this model is estimated and the data are adjusted accordingly (Hoefer et al 2007). The major difficulty in the US model is to calculate the number of legal immigrants (legally resident foreign born) from the sum of the constituent immigration components as there is no current population registry available in the United States.

30. The use of "residual" estimation methods based on the differences between census data and other registries of immigrants is not common in European states, mostly because the censuses are thought to be seriously undercounting irregular migrants in Europe (cf. Lederer p.197). The two major exceptions are the UK and Spain where the residual method has recently been applied for the estimation of the size of the illegally resident population.

31. In the UK, the Home Office commissioned a report that provides an estimate of the foreign-born population living there illegally based on a residual method that is summarized in Woodbridge (2005). The methodology involves the same calculations as in the US case above but adds another category of migrants who are in the UK on a "quasi-legal basis" (Q), for example asylum-seekers or refused applicants with leave to remain. Thus, the total foreign-born population can be calculated as

$$(1)\ FB = (L - (M_L+E_L)) + (T - (M_T+E_T)) + (Q - (M_Q+E_Q)) + R$$

with M_L, E_L … denoting group-specific mortality and emigration rates.

32. The total foreign-born population is obtained from the census of 2001, thus the estimate of the residual R pertains to April 2001. As no explicit undercount estimates are available for the foreign-born population FB, a range of undercounts (lower-middle-upper) is applied that results in a range of estimates of R.[14] An additional complication and uncertainty is given by the fact that in the UK no registered stock figure for legal permanent immigrants L is available – the stock is thus obtained by adding the number of migrants granted settlement each year since 1970 adjusted for likely emigration and death.[15]

[14] The central estimate of the unauthorized resident population in the UK in April 2001 was 430,000 while the lower and upper range estimates were 310,000 and 570,000, respectively.

[15] The number of legal immigrants granted settlement in the UK before 1970 was estimated with an alternative method.

2. Simple Comparison of Registers

33. Another exception to the general assumption that illegally resident foreigners are not well covered in the census or other official registers is provided by Spain. Here the law allows foreigners – whatever their legal status – to be included in the municipal register if they can prove residence in that municipality (Aparicio and Ruiz 2008). Irregular migrants, particularly since 2001, have been encouraged to register to obtain health benefits and because they can use this as proof of residence for later regularization while there are hardly any disincentives connected to registering as these data are not utilized for removing unauthorized residents from the country. Consequently, the difference between the figures in the municipal population register and those of the register of foreigners with residence permits has been used as an indicator for the number of irregular foreigners in the country. Table 3 provides an overview of the difference between these two registries from 1998-2005.

Table 3: Comparison of the Figures of the Foreign Population in the Municipal Register with the Figures of the Register of Foreigners with Resident Permits (1998-2005)

Register	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**
Municipal Register	637.085	748.954	923.879	1.370.657	1.977.946	2.664.168	3.034.326	3.730.610
Foreigners with Resident Permit	719.647	801.332	895.720	1.109.060	1.324.001	1.647.011	1.977.291	2.738.932
Difference (MR – Foreigners with Resident Permit	- 82.562	-52.378	28.159	261.597	653.945	1.017.157	1.057.035	991.678

Source: Instituto Nacional de Estadística and Ministerio de Trabajo y Asuntos Sociales, Observatorio Permanente de la Inmigración,

34. As can be seen from Table 3, the difference between the two registers has rapidly increased up to 2004, when more than a million foreigners more were registered in the municipal register than had a residence permit in Spain. This compares with about 690,000 foreigners in an irregular situation who had applied for regularization in the latest regularization campaign in 2005 (see Section 3.3.1 above).

35. Despite the apparent usefulness of comparing the two registers, there are also several problems connected with using this method. First, the municipal population register itself has many flaws which render the number obtained somewhat unreliable, particularly with regard to the number of registered persons who have already left the country but did not de-register (as commonly observed in population registers). Second, the number of foreigners with resident permits is not necessarily the same as the number of legal foreigners in the country (due to emigration, death, naturalization, etc.). And third, the simple comparison of registers does not allow the estimation of foreigners in the country who are neither in one nor the other register recorded as this would require an individual matching of records (see Section 3.1.4).

B. COMPARISON OF IMMIGRATION AND EMIGRATION STATISTICS

36. In an overview of some non-European studies on the size of illegal migrant populations Lederer (2004, p. 192ff) describes several studies that attempted to estimate the number of undocumented migrants in the US from Mexico based on demographic data from Mexico. One of the methods applied used the Mexican census results (1960 and 1970) to calculate the difference between the actual and the expected population living in Mexico as an approximation of actual emigration. Assuming that most emigration is to the US and subtracting legal emigration, the authors arrive at an estimate of undocumented Mexicans in the US. Another method used observed and expected sex-ratios in Mexico to estimate the number of undocumented migrants in the US ("missing males").

37. More recently, attempts were made to apply the method in the European context. Delaunay and Tapinos (1998a, pp. 42-56) detail their own calculations on emigration data from Morocco and Tunisia to estimate the number of emigrants from these country living illegally in the main European destination countries. The authors use both data from Moroccan censuses in 1982 and 1994 and Tunisian censuses in 1984 and 1994 and also test the sex-ratio method on the Moroccan case. However, the authors admit that their results for Morocco[16] suffer from serious data deficiencies in the census data available and that the estimation for Tunisia was not possible due to a lack of reliable data.

38. These observations are in line with general observations on census and emigration data in European countries of origin: emigration data are undercounting and census results are often unreliable. Moreover, the method is complicated by the fact that – unlike the Mexican-US case – migrants originate from a large number of countries and disperse across a larger number of receiving states in Europe. However, despite these general caveats, emigration data could still be a valuable complementary source of information on irregular migrants in European countries. For example, a study on emigration from Ukraine presents estimates on the number of Ukrainian migrants abroad (often illegally), based on information received from embassies and consulates in the reception countries (Malynovska 2004).[17] Comparing this information with the number of legal immigrants from Ukraine registered in the most important destination countries could be an alternative indicator for estimating the number of illegal residents from certain countries of origin and destination.

[16] The clandestine population from Morocco in Europe was estimated by the expected population method to be around 370,000 and by the sex-ratio method to be between 600,00 and 700,000 persons.

[17] "According to the information from Ukrainian embassies, Ukrainian labor migration has the following structure in terms of countries of destination: in Poland there are 300 thousand labor migrants, in Italy and the Czech Republic, 200,000 (each), in Portugal, 115,000, In Spain, 100,000, in Turkey, 35,000, in the USA, 20,000. The number of Ukrainians who work in the Russian Federation is estimated to be 1 million people." (Malynovska 2004, p.14)

V. CONCLUSIONS

39. This paper has attempted to describe some existing and applicable methods to the estimation of irregular migration phenomena in North America and the European Union and its neghbouring countries. The emphasis was on methodological clarification and categorisation of statistically grounded estimation methods and not on a full coverage and evaluation of existing estimates.[18] The basic logic of methodologies was laid out in a commonly accessible language leaving statistical intricacies and mathematical formulas out in the references for access by the specialists. A short discussion of special data requirements or data problems was included as a guide to the description, categorisation, evaluation and classification of existing estimates found in European countries.

40. In the conclusions of their comprehensive review of existing estimation methodologies in Europe, Delaunay and Tapinos (1998, p.72) had stated that "the most obvious and disappointing finding which emerges … is that we have practically nothing that is well-founded". The next review of methods by Pinkerton et al (2004) examined the applicability of estimation methods to the UK case and sounded already much more optimistic, resulting eventually in the first official estimate on the UK (Woodbridge 2005 uses the residual method). The classification scheme presented in Table 1, only 10 years after the first one by Delaunay and Tapinos (1998), could add again a number of new estimation methods and techniques that may or may not be fruitfully applied in the different contexts. Thus, while we do not yet have generally accepted estimation methods for the specific regions, we have already made some progress along the way.

REFERENCES

Belser, P., de Cock, M. and Ferhard M. (2005), ILO Minimum Estimate of Forced Labour in the World, ILO, Geneva, April 2005

Chelimsky, E. (1991), Politics, Policy Making, Data, and the Homeless, in: Housing Policy Debate, Vol. 2, Issue 3, pp. 683-697

Delaunay, D. and Tapinos, G. (1998a), "La mesure de la migration clandestine en Europe." Volume 1: rapport de synthese, Eurostat Working Papers, Series "Population et conditions sociales", 3/1998/E/No7

Delaunay, D. and Tapinos, G. (1998b), "La mesure de la migration clandestine en Europe." Volume 2: rapport des experts, Eurostat Working Papers, Series "Population et conditions sociales", 3/1998/E/No7

Hoefer, M., Rytina, N. and Campbell, C. (2007), Estimates of the Unauthorized Immigrant Population Residing in the United Sates: January 2006, Office of Immigration Statistics, Policy

[18] A full coverage, evaluation and classification of existing estimates for a number of European countries is the task of the CLANDESTINO project (EU-FP 6 Project, 2007-2009) in which a full methodological classification scheme is elaborated.

Directorate, U.S. Department of Homeland Security, available at: http://www.dhs.gov/xlibrary/assets/statistics/publications/ill_pe_2006.pdf

Jandl, M., (2004), The Estimation of Illegal Migration in Europe, Studi Emigrazione/ Migration Studies, vol. XLI, No. 153, March 2004, pp. 141-155

Jandl, M. and Kraler, A. (2006), Links Between Legal and Illegal Migration, in: Poulain, M., Perrin, N. and Singleton, A. (2006), THESIM – Towards Harmonized European Statistics on International Migration, UCL Presses, Louvain-La-Neuve 2006, pp. 337-371

Lederer, H. W. (2004), Indikatoren der Migration. Zur Messung des Umfangs und der Arten von Migration in Deutschland unter besonderer Berücksichtigung des Ehegatten- und Familiennachzugs sowie der illegalen Migration, efms, Bamberg 2004

Leerkes, A., Engbersen G. and van San, M. (2007), Shadow Places: Patterns of Spatial Concentration and Incorporation of Irregular Immigrants in the Netherlands, in: Urban Studies, Vol. 44, No. 8, 1–26, July 2007

Leun, J. van der, G. Engbersen en P. van der Heijden (1998): Illegaliteit en criminaliteit: schattingen, aanhoudingen en uitzettingen. Onderzoeksrapport, Faculteit der Sociale Wetenschappen/Sociologie, Erasmus Universiteit, Rotterdam.

Malynovska, O. (2004), International migration in contemporary Ukraine: trends and policy, in: Global Migration Perspectives, No.14, October 2004, available at: http://www.gcim.org/attachements/GMP%20No%2014.pdf

Natale, M. (1998), Determining irregular foreigners in the Italian population, in: Delaunay, D. and Tapinos, G. (1998b), "La mesure de la migration clandestine en Europe." Volume 2: rapport des experts, Eurostat Working Papers, Series "Population et conditions sociales", 3/1998/E/No7

Petersen, C. (1896), The Yearly Immigration of Young Plaice into the Limfjord from the German Sea, in: Report of the Danish Biological Station to the Ministry of Fisheries, Bd.6, S.1-48)

Pinkerton, C.G., McLaughlan and Salt, J. (2004), Sizing the illegally resident population in the UK, Home Office Online Report 58/04, available at: http://www.homeoffice.gov.uk/rds/pdfs04/rdsolr5804.pdf

Sciortino, G. (2003), Regularization of Foreigners in Italy, unpublished manuscript, Communication to Michael Jandl on 3.4.2003

Sikkel, D., van der Heijden, P. and van Gils, G. (2006), Summary, in: *Methoden voor omvangschattingen van verborgen populaties, met name illegalen*, WODC rapport 243, Meppel: Boom Juridische Uitgevers

United Nations Office on Drugs and Crime, Methodology, in: World Drug Report 2007, Vienna, p.257-274

Van der Heijden, P. (2006), Estimating the size of the illegal population in the Netherlands, Graduate School Lecture, University of Utrecht, 9 October 2006

Van der Heijden, P., Bustami, R., Cruijff, M. , Engbersen, G. and van Houwelingen, H. (2003), Point and interval estimation of the truncated Poisson regression model, in: Statistical Modelling, 3, pp. 305-322.

Van der Heijden, P., Cruijff, M. and van Houwelingen, H. (2003), Estimating the size of a criminal population from police registrations using the truncated Poisson regression model, in: Statistica Neerlandica, 57, pp. 289-304.

Van der Heijden, P., van Gils, G., Cruijff, M. and Hessen, D. (2006), Summary of 'An estimate of the number of illegal aliens residing in the Netherlands in 2005', October 2006

Van der Leun, J., G. Engbersen en Peter van der Heijden (1998). *Illegaliteit en criminaliteit: schattingen, aanhoudingen en uitzettingen*. Rotterdam: Erasmusuniversiteit

Vogel, D. (2002): Ausländer ohne Aufenthaltsstatus in Deutschland. Methoden zur Schätzung ihrer Zahl. In: Minthe, Eric (ed.): Illegale Migration und Schleusungskriminalität. Wiesbaden, Eigenverlag der Kriminologischen Zentralstelle e.V., 2002, pp. 65-78.

Zhang, L. (2007), On the estimation of illegal residents in Norway (Draft report), Statistics Norway, unpublished manuscript, Oslo 2007

Zhang, L. (2008), On the estimation of illegal residents in Norway (Final report), Statistics Norway, Oslo 2008, forthcoming

*****