

Distr.: General  
24 October 2017

English

---

## Европейская экономическая комиссия ООН

### Конференция европейских статистиков

#### Рабочая сессия по статистике миграции

Женева, Швейцария

30-31 октября 2017 г.

Пункт 3 предварительной повестки дня

#### Интеграция данных и административные данные

## **Оценка миграции на основе данных налоговой отчетности Португалии в сочетании с другими административными регистрами и методами оценки**

Подготовлено Статистическим управлением Португалии\*

### *Резюме*

Статистическое управление Португалии в настоящее время изучает возможность использования административных данных при производстве официальной статистики о населении на основе целого ряда источников, которые были первоначально задействованы с целью разработки пилотной модели подсчета численности населения с использованием методологии определения «признаков жизни», но которые в равной мере целесообразны и для других целей, например, для усовершенствования статистики миграции. Налоговый регистр является одним из самых важных и комплексных источников данных, который должен быть рассмотрен на данном этапе, поскольку он предоставляет информацию по гражданам и иностранным лицам всех возрастов, проживающим в Португалии или за ее пределами. Для измерения миграционных потоков будет использован переходный метод. После необходимых поправок (в том числе калибровки весов), полученные таким образом результаты будут далее использованы для подсчета коэффициента распределения по субнациональным районам (фактическое количество иммигрантов/эмигрантов на национальном уровне будет оцениваться в соответствии с текущими методами). Рассматриваемое исследование также охватит потоки внутренней миграции посредством

\*Подготовлено г-жой Паулой Круз

сравнения почтовых индексов налогоплательщиков, присутствующих в стране на протяжении двух лет подряд. Несмотря на ограниченность имеющихся данных, полученные в настоящее время результаты свидетельствуют о возможном использовании налоговых регистров при составлении оценок миграции. На первом этапе, однако, этот процесс будет ограничиваться повышением качества имеющейся миграционной статистики, в частности в отношении составления оценок на более детализированном географическом уровне. Более далеко идущий подход потребует тесного сотрудничества с налоговыми и таможенными органами Португалии для оптимизации использования налоговых регистров в статистических целях.

## I. Введение

1. Статистическое управление Португалии в настоящее время изучает возможность использования административных данных в производстве официальной статистики о населении, в то числе: национального гражданского регистра населения, регистра иностранного населения, регистра образования, налогового регистра, регистра социального обеспечения, регистра занятости и безработицы и т.д.

2. Эти источники данных первоначально были использованы для разработки пилотной модели подсчета численности населения на основе методологии «*признаков жизни*», но они в равной мере целесообразны и для использования в других целях, например, повышения качества статистики миграции, благодаря наличию информации о резидентстве на протяжении двух лет подряд.

3. Хотелось бы напомнить, что в настоящее время годовые оценки численности иммигрантов и эмигрантов составляются Статистическим управлением Португалии на основе данных обследования рабочей силы (ОРС) в сочетании с данными регистров иностранных граждан в отношении иммиграционных потоков и данных обследования эмиграции (модуль в рамках ОРС) в отношении эмиграционных потоков, которые доступны только на национальном уровне из-за недостаточной надежности имеющихся оценок на меньшем уровне агрегирования.

4. Согласно уже проведенному анализу регистр налогоплательщиков (в том числе заполнители, лица, освобожденные от заполнения и иждивенцы) Налоговый регистр является одним из самых важных и комплексных источников данных, который должен быть рассмотрен при составлении субнациональных оценок, так как он содержит информацию на уровне почтовых индексов по гражданам и иностранным лицам всех возрастов (с 2011 года регистрация новорожденных является обязательной), проживающим в Португалии или за ее пределами.

5. В соответствии со Статьей 16 Кодекса подоходного налогообложения физических лиц, физическое лицо является резидентом Португалии, если оно проживает более 183 дней на территории Португалии в течение 12-месяцев, начинающихся или заканчивающихся в соответствующем календарном году. В случае проживания в течение более короткого периода, физическое лицо

также считается резидентом Португалии, если в любое время в этот период оно имеет жильем с намерением сохранить его в качестве постоянного места проживания.

6. Физическое лицо, если оно не докажет обратное, также считается резидентом Португалии, если оно переезжает для проживания на другую территорию исключительно с целью получения выгоды от явно более благоприятного налогового режима, включенного в одобренный правительством перечень.

7. Такая концепция налогового резидентства, вступившая в силу 1 января 2015 года, более тесно связана с периодом эффективного резидентства физического лица в стране, чем предыдущая. Так до 31 декабря 2014 года, в случае если физическое лицо проживало, например, 9 месяцев на территории Португалии, оно будет считаться резидентом Португалии в отношении всего года; согласно новому положению, оно может считаться резидентом Португалии в первые 9 месяцев и нерезидентом в оставшееся время в течение года [1, 2].

8. Поскольку имеющиеся данные охватывают период с 2014 года по 2016 год, будет интересно проанализировать последствия этих концептуальных изменений на относительное распределение потенциальных мигрантов на субнациональном уровне.

9. Цель данного документа заключается в описании методологии, которая будет принята в рамках рассматриваемого исследования, наряду с представлением краткого описания используемых источников данных, включая уже выявленные вопросы качества и охвата. Также будут представлены некоторые предварительные результаты, касающиеся региональной разбивки миграционных потоков, а также основные ограничения и выводы осуществляемого исследования.

## II. Описание источников данных

10. Основные источники данных состоят из регистров налогоплательщиков (в том числе заполнители, лица, освобожденные от заполнения и иждивенцы) за 2014, 2015 и 2016 годы (отчетной датой является 31 декабря каждого года), которые содержат следующее общее количество записей:

Таблица 1 – Налоговые регистры

	2014	2015	2016
Страна проживания = Португалия	14 029 321	14 022 860	13 958 720
Страна проживания = другая	1 005 260	1 703 751	1 610 837

11. Необходимо отметить, что последний выпуск данных за 2016 год обусловил более ограниченный анализ рассматриваемой информации.

12. Эти файлы содержат базовую демографическую и географическую информацию о налогоплательщике со 100% охватом данных, в частности:

- пол;
- место рождения;
- страна гражданства (Португалия/другая);
- место рождения (район, муниципалитет и округ);
- адрес (населенный пункт, почтовый индекс);
- страна проживания (Португалия истинно или ложно).

13. Данные налогового регистра демонстрируют очевидный избыточный охват при сравнении с годовыми оценками постоянного населения (в среднем 10 341 908 человек в период 2014-2016 гг.). Вместе с тем, как показано на графиках ниже, налогоплательщики, определенные в качестве резидентов Португалии, показывают аналогичное сравнительное распределение по половому признаку; что касается возраста, то за исключением первой группы (0-4 года), все возрастные группы показывают избыточный охват в налоговом регистре, в частности в группе 85 лет и старше. Это обусловлено тем фактом, что умершие налогоплательщики, записи о которых никогда не удаляются из регистра в фискальных целях, а сохраняются в виде неактивных налогоплательщиков в поле «статус», не исключаются, что было бы необходимо для правильной идентификации таких случаев. Вместе с тем необходимо отметить, что это упущение не мешает проведению исследованию миграционных потоков, поскольку они подразумевают изменения в статусе резидентства лица, и соответственно, менее подвержены тому, чтобы быть связанным с умершими или неактивными налогоплательщиками.

14. Другой важный вывод касается наличия большого сходства между данными налоговых регистров и сравнительным распределением оценок постоянного населения на уровне региона NUTS 2 в рассматриваемый период.

15. Ожидается, что в будущих выпусках этих файлов будут предоставлены дополнительные данные, в частности в отношении:

- Событий, которые оказали влияние на объекты регистра, например, дата последнего обновления;
- Ежеквартальных выпусков информации, которые позволят более эффективно определять тип миграционного перемещения (временное или постоянное).

16. Наконец, как уже отмечалось ранее, другие источники данных также будут рассматриваться с целью решения проблемы избыточности охвата на основе применения методологии «*признаки жизни*», которая главным образом направлена на обнаружение лиц, являющихся активными, по крайней мере, в двух административных регистрах. Для этого будет использована уже имеющаяся пилотная модель подсчета численности населения на основе ряда административных регистров касательно 2015 года[3].

График 1 – Постоянное население в разбивке по полу

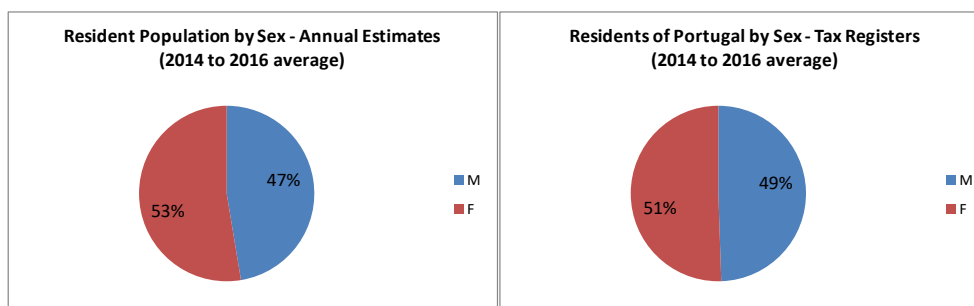


График 2 – Постоянное население в разбивке по возрастным группам

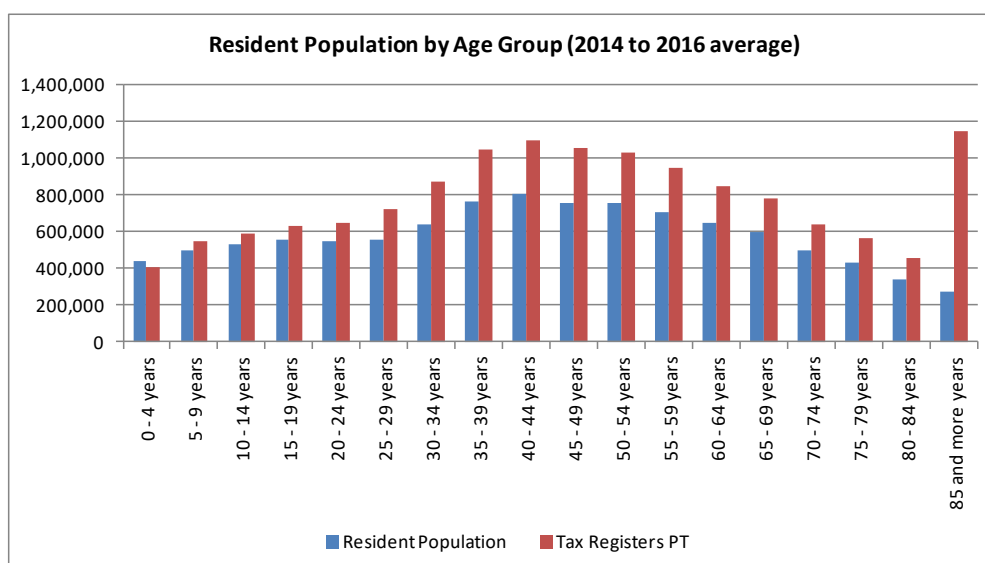


Таблица 2 – Сравнительное распределение по NUTS 2: оценки постоянного населения (ResPop) и данные налоговых регистров (TaxReg)

D_NUTS 2	2014		2015		2016	
	ResPop	TaxReg	ResPop	TaxReg	ResPop	TaxReg
Norte	34.91%	33.91%	34.85%	33.79%	34.77%	33.71%
Centro	21.82%	22.10%	21.82%	21.94%	21.76%	21.79%
Lisboa	27.08%	26.46%	27.20%	26.72%	27.37%	26.94%
Alentejo	7.07%	6.93%	7.00%	6.92%	6.97%	6.88%
Algarve	4.26%	5.01%	4.27%	5.03%	4.28%	5.07%
RAA*	2.37%	2.70%	2.38%	2.72%	2.38%	2.73%
RAM*	2.49%	2.89%	2.48%	2.88%	2.47%	2.88%

\*RAA – Região Autónoma dos Açores; RAM – Região Autónoma da Madeira.

### III. Обзор методологии

17. Предлагаемая методология будет близко следовать методологии, используемой Статистическим управлением Канады, как изложено в ряде

документов, датируемых периодом 1998 - 2016 годами после последующих изменений [4, 5]. Основная идея будет заключаться в том, чтобы использовать налоговые регистры в качестве источника подробной географической информации касательно смены резидентства налогоплательщика в Португалии на другую страну и наоборот (международная миграция) или в пределах Португалии (внутренняя миграция).

## A. Международная миграция

18. Прежде всего, необходимо отметить, что из-за проблемы избыточного охвата и отсутствия соответствующей метаинформации, связанной с налоговым регистром, фактическое количество иммигрантов/эмигрантов на национальном уровне будет по-прежнему оцениваться с использованием текущих методов.

19. *Переходный* метод, в соответствии с которым население страны сравнивается в два момента времени ( $t$  и  $t+1$ ), вместо одного конкретного периода времени ( $t$ ,  $t+1$ ) при подходе, основанном на событиях [6], будет использован для идентификации налогоплательщиков-мигрантов, то есть, лицо будет считаться:

- Потенциальным эмигрантом, если он/она проживает в стране в момент времени  $t$ , но отсутствует в момент времени  $t+1$  (за исключением смерти) и
- Потенциальным иммигрантом, если он/она отсутствует в момент  $t$ , но присутствует в стране в  $t+1$  (за исключением новорожденных).

20. В первом случае будут рассматриваться только налогоплательщики, зарегистрированные на протяжении двух последовательных года с изменением переменной, обозначающей является ли налогоплательщик резидентом в Португалии (от «Истинно» до «Ложно») (это обновление само по себе, в принципе, исключает возможную смерть). На этом этапе проекта не представляется возможным учесть не заявленные изменения в статусе проживания.

21. Иммигранты, с другой стороны, будут определены одним из двух способов:

- Налогоплательщики, зарегистрированные в оба года с изменением переменной, обозначающей является ли налогоплательщик резидентом в Португалии (от «Истинно» до «Ложно»);
- Налогоплательщики, зарегистрированные только во второй отчетный год и рожденные до этого года.

22. Были получены некоторые предварительные результаты, которые подтверждают прогнозируемый избыточный охват в абсолютных совокупных величинах при сравнении с существующими оценками международных мигрантов (2015).

23. Посредством использования дополнительных источников административных данных и зашифрованный для каждого лица номер плательщика НДС на основе упомянутой ранее методологии «*признаки жизни*», в то же время, стало возможным сократить эти величины следующим образом:

Таблица 3 – Потенциальные международные мигранты на основе налоговых регистров

	Налоговый регистр (2015)		Оценки международных мигрантов (Статистическое управление Португалии, 2015 г.)
Потенциальные эмигранты	226 558	Эмигранты (всего)	101 203
Потенциальные иммигранты	31 519	Иммигранты (постоянные)	29 896

24. Полученные данные о потоках опять же указывают на избыточный охват, в частности в отношении общего количества эмигрантов, что может быть вызвано следующим:

- Включение потенциальной эмиграции продолжительностью ниже порогового значения в 3 месяца;
- Изменения статуса резидентства, связанные с новой концепцией, а не с эффективной миграцией;
- Несоответствие между переменными, относящимися к одной и той же области (например, резидентство).

25. Дальнейшие поправки избыточного охвата, таким образом, могут быть сделаны на основе дополнительных административных источников и правил обеспечения согласованности. Вместе с тем, из-за отсутствия соответствующей информации на данном этапе исследования может быть осуществлен только этот последний тип проверки данных, в частности в отношении данных, связанных с резидентством.

26. После внесения соответствующих корректировок и на основе существующих оценок<sup>1</sup> по полу, возрасту, гражданству (Португалия/другое) и/или типу миграции, будут использованы процедуры калибровки для корректировки веса каждого регистра (до корректировки все веса равны 1; после корректировки веса для различных категорий, в отношении которых отмечается избыточный охват, равны менее 1) детально следуя предлагаемой [7] методологии.

<sup>1</sup> Микроданные обследования эмиграции и официальные оценки постоянной иммиграции.

27. Полученные таким образом результаты будут далее использованы для расчета коэффициента распределения по субнациональным районам (а именно НАЕС 2), в основе которого будут лежать административные деления, соответствующие действительным почтовым кодам. Эти коды будут определены используя последние файлы, предоставленные почтовой службой Португалии. Почтовые коды, соответствующие почтовым ящикам, будут исключены; почтовые коды, связанные с более чем одним административным делением будут условно причислены к самому значимому, на основе соответствующего критерия.

28. В случае эмиграционных потоков, страны назначения также будут проанализированы на основе зарегистрированного статуса резидентства в отношении 2015 года. Вместе с тем необходимо отметить, что поле, о котором идет речь, представляет собой текстовое поле без определенного формата. Существующие регистры содержат информацию о странах, регионах, городах, наименованиях улиц и/или почтовых индексах, которая вводится вручную на различных языках. В этой связи использование таких данных предполагает длительный процесс выверки данных с целью сопоставления полученной в результате строки со стандартизированными регистрами стран, городов и почтовых индексов.

29. Основным программным обеспечением, которое будет использоваться для извлечения, предварительной обработки и анализа имеющихся данных, является R Studio и следующие пакеты и библиотеки: ROracle, карты, ggmaps, plyr, sqldf, data.table, stringr, dplyr, tm и DescTools. Также, для геокодирования будут использованы Google API.

## **В. Внутренняя миграция**

30. Идентификация внутренних мигрантов-налогоплательщиков будет основана на сопоставлении почтовых индексов налогоплательщиков, которые проживают в стране (PT="V") на протяжении двух лет подряд, в частности 2014 и 2015 годах, и одновременно демонстрируют «признаки жизни» в последний год.

31. Полученные данные будут проанализированы на более агрегированном уровне (а именно NUTS 2) на основе метода анализа социальной сети (SNA) [8] и программного обеспечения Gephi [9, 10], доступного на сайте <http://gephi.org>.

## **IV. Некоторые предварительные результаты**

### **A. Международная миграция**

#### **1. Эмиграционные потоки**

32. Некалиброванный регистр потенциальных эмигрантов насчитывает 194 533 человека, то есть почти в два раза больше официальных оценок (101 203 человека), что может быть обусловлено изложенными выше причинами, главным образом лежащей в основе налоговых регистров концепции



резидентства Португалии, которая была принята в 2014 году. Так, анализ потенциальной эмиграции в период с 2015 по 2016 годы, проведенный на основе новой концепции резидентства, демонстрируют значительно меньшую разницу с официальными оценками: 127 561 и 97 151 человек, что также согласуется с наблюдаемым сокращением этих видов миграционных потоков.

33. Переменные, которые были использованы для подсчета калиброванных весов каждого потенциального эмигранта, не содержат отсутствующих значений и включают следующее: возраст, пол и страна гражданства. Были осуществлены две процедуры калибровки: (1) первая была основана на официальных оценках по возрастной группе, полу и национальности (групп стран); (2) вторая – по полу, в разбивке по функциональным возрастным группам (0-14,15-24,25-64,65+) на основе имеющихся микроданных (обследование эмиграции).

34. Полученные коэффициенты распределения по региону предыдущего резидентства НАЕС 2 представлены в таблице ниже, наряду с показателями, основанными на некалиброванных данных наблюдения:

35.

Таблица 4 – коэффициенты распределения по региону предыдущего резидентства НАЕС 2 (потенциальные эмигранты, 2015)

Регион НАЕС 2	Калиброванные данные		Некалиброванные наблюдения
	(1)	(2)	
Norte	40.30%	39.19%	38.92%
Centro	25.24%	25.40%	25.96%
Área Metropolitana de Lisboa	23.43%	23.65%	22.98%
Alentejo	3.53%	3.54%	3.57%
Algarve	3.89%	4.89%	5.36%
Região Autónoma dos Açores	0.34%	0.31%	0.29%
Região Autónoma da Madeira	3.27%	3.02%	2.92%

36. Можно отметить, что все три распределения крайне схожи, о чем свидетельствует высокий коэффициент корреляции (примерно 99% во всех случаях). Они также тесно коррелируют (примерно 98%) с коэффициентом распределения постоянного населения в 2015 году и с распределением, полученным на основе данных эмиграционного обследования.

37. Эти результаты являются верными в отношении изменений статуса резидентства в период с 2015 по 2016 год, указывая на очевидно незначительное воздействие новой концепции налогового резидентства на региональное распределение соответствующих лиц:

Таблица 5 – Коэффициенты распределения по региону НАЕС 2 предыдущего места проживания (потенциальные эмигранты, 2016)

Регион НАЕС 2	Некалиброванные данные
---------------	------------------------

Norte	40.73%
Centro	26.97%
Área Metropolitana de Lisboa	20.76%
Alentejo	3.39%
Algarve	4.82%
Região Autónoma dos Açores	0.76%
Região Autónoma da Madeira	2.57%

38. Что касается основных стран назначения, полученные результаты, похоже, совпадают с данными, полученными из регистров консульств Португалии [11], за исключением Швейцарии и Люксембурга, как показано в Таблице 6. Вместе с тем, необходимо отметить, что почти 18% рассматриваемых регистров состоят из невыясненных случаев из-за недостаточной информации.

Таблица 6– Рейтинг основных стран назначения (Европа)

Страна	Доклад об эмиграции (2015)	Налоговые регистры (2015)
Соединенное Королевство	1	2
Франция*	2	1
Швейцария	3	7
Германия	4	4
Испания	5	5
Бельгия*	6	6
Люксембург	7	3

\*Данные доклада об эмиграции относятся к 2013 году.

## 2. Иммиграционные потоки

39. Как представлено выше, показатели численности потенциальных иммигрантов, полученные на основе налоговых регистров, весьма близки к официальным оценкам, относящимся к тому же периоду (2015 г.). Необходимо отметить, что этот результат был получен преимущественно за счет использования методологии «*признаки жизни*», что позволило сократить изначальное количество наблюдений. Как и в случае с эмиграцией, это может быть обусловлено внедрением новой концепции, принятой в 2014 году.

40. После проведения процедуры калибровки на основе переменной пола в разбивке по 5-летним возрастным группам и национальности (группы стран), были получены следующие коэффициенты распределения резидентства на уровне региона НАЕС 2:

Таблица 7 – Коэффициенты распределения по региону назначения НАЕС 2 (потенциальные эмигранты, 2015)

Регион НАЕС 2	Калиброванные наблюдения	Некалиброванные наблюдения
Norte	20.84%	20.09%
Centro	14.68%	15.43%
Área Metropolitana de Lisboa	46.79%	48.90%
Alentejo	4.48%	4.42%
Algarve	10.67%	8.48%
Região Autónoma dos Açores	1.13%	1.21%
Região Autónoma da Madeira	1.41%	1.47%

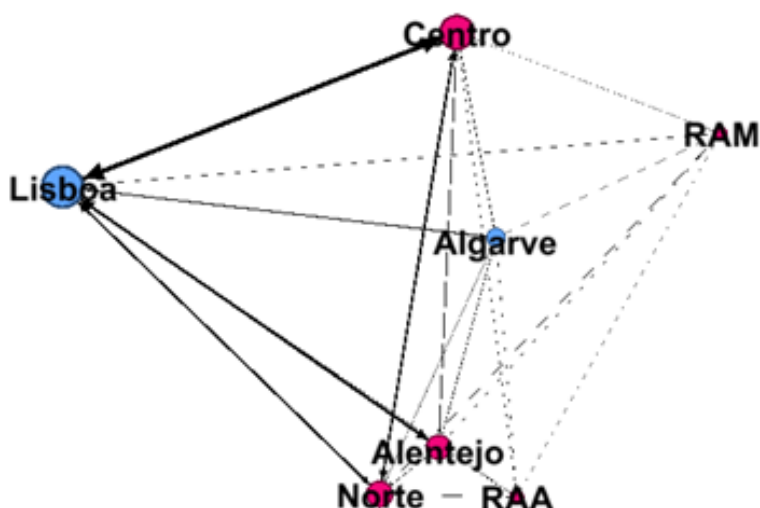
41. Как и в случае с эмиграцией, важно отметить, что эти два распределения являются достаточно схожими, о чем свидетельствует высокий коэффициент корреляции (примерно 99%, во всех случаях). Они также тесно коррелируют (примерно 90%) с коэффициентом распределения, полученным из ОРС и коррелируют в меньшей степени с соответствующим распределением постоянного населения (75%).

## В. Внутренняя миграция

42. Предварительные результаты указывают на наличие 104 216 внутренних эмигрантов по региону НАЕС 2 в рассматриваемый период (2014-2015 гг.).

43. На Графике ниже красным обозначены регионы с отрицательной чистой миграцией, а размер каждой точки отражает объем мигрантов, что позволяет отметить, что только регионы НАЕС 2 Лиссабона и Алгарве демонстрируют больший приток, чем отток мигрантов в период 2014-2015 гг. Эта закономерность в какой-то мере схожа с закономерностью, обуславливающей распределение потенциальных иммигрантов региона НАЕС 2 (Таблица 7).

График 3 – Потенциальные внутренние мигранты по НАЕС 2 – SNA



## V. Заключение

44. Имеющиеся в настоящее время налоговые данные представляют определенные ограничения, самое важное из которых относится к представленному на сегодняшний день отчетному периоду регистров (31 декабря каждого года). Ежеквартальные обновления позволят более эффективно определить объем миграционных потоков, поскольку станет возможным измерить фактический период резидентства лиц, которые меняют данные о своем резидентстве один или несколько раз на протяжении года. В конечном итоге использование предлагаемой методологии позволит определить также потенциальную краткосрочную круговую миграцию.

45. Несмотря на ограниченность имеющихся данных, полученные в настоящее время результаты свидетельствуют о возможном использовании налоговых регистров при составлении оценок миграции. На первом этапе, однако, этот процесс будет ограничиваться повышением качества имеющейся миграционной статистики, в частности в отношении составления оценок на более детализированном географическом уровне, в соответствии с ранее представленной методологией (согласно которой фактическое количество иммигрантов/эмигрантов на национальном уровне будет оцениваться в соответствии с текущими методами)

46. Более далеко идущий подход потребует тесного сотрудничества с налоговыми и таможенными органами Португалии для оптимизации использования налоговых регистров в статистических целях.

## VI. Библиография

- [1] Nogueira D. (2015): Tributação das pessoas singulares: residentes, não residentes e residentes não habituais. Tese de Mestrado em Direito (Direito Fiscal). Universidade Católica Portuguesa.
- [2] Flores C. (2016): O Elemento da Residência Fiscal para as Pessoas Singulares. Tese de Mestrado em Direito. Universidade do Porto.
- [3] Gabinete para os Censos 2021 (2015): Estudo de viabilidade para os Censos 2021 - Linhas gerais do novo modelo para os Censos 2021, a testar em 2016 (Relatório QUAR). INE.
- [4] Demography Division (2016): Population and Family Estimation Methods at Statistics Canada. Statistics Canada.
- [5] Small Area and Administrative Data Division (1998): Description of the methodology used to create migration data from taxation records. Statistics Canada.
- [6] Wisniewski A. (2017): Combining Labour Force Survey data to estimate migration flows: The case of migration from Poland to the UK. *Journal of the Royal Statistical Society. Series A: Statistics in Society*.
- [7] Wallgren A. and Wallgren B. (2014): Register-Based Statistics - Administrative Data for Statistical Purposes. John Wiley & Sons, Ltd.
- [8] Maier G, Vyborny M, 2008, "Internal migration between US States: A social network analysis", in *Migration and Human Capital* Eds Poot J, Waldorf B, Wissen L W (Edward Elgar, Cheltenham, Glos) pp 75–93.
- [9] Bastian M., Heymann S., Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media*.
- [10] Jacomy M, Venturini T, Heymann S, Bastian M (2014) ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLoS ONE* 9(6): e98679. <https://doi.org/10.1371/journal.pone.0098679>.
- [11] Portal das Comunidades Portuguesas (2015): Relatório da Emigração 2015. Gabinete do Secretário de Estado das Comunidades Portuguesas, pp 49.