The Economic and Social Review, Vol. 46, No. 4, Winter, 2015, pp. 567–603

To Weight or Not To Weight?
A Statistical Analysis of How Weights Affect the Reliability of the *Quarterly National Household Survey* for Immigration Research in Ireland

NANCY DUONG NGUYEN
University College Dublin

PATRICK MURPHY
University College Dublin

Abstract: Ireland has an immigrant population that is proportionally among the largest in the EU. Considering the impact this has had on Irish society, a surprisingly small amount of quantitative research has been performed to date. This paper evaluates the effectiveness of the weighting schemes associated with the *Quarterly National Household Survey* (*QNHS*), the primary data source for immigration research in Ireland. This is the first time that the *QNHS* weighting mechanism has been formally evaluated in the literature. Our analysis shows that there are significant issues relating to the weighting mechanism. This has major consequences for quantitative research on immigration in Ireland.

I AIMS AND BACKGROUND

Large-scale immigration is a relatively new phenomenon in Ireland. This leads to a lack of official data available on the country's immigrant population. Researchers in the field have to rely on a small number of data sources such as the *Census of Population*, the *Quarterly National Household Survey* (*QNHS*), the *European Union Statistics on Income and Living*

E-mail: duong.nguyen@ucdconnect.ie, patrick.murphy@ucd.ie

Conditions (*EU-SILC*) and the *National Employment Survey* (*NES*). Of these the *QNHS* has proved to be the most used resource for researchers. Conducted by the Irish Central Statistics Office (CSO), the *QNHS* is a voluntary sample survey which makes it prone to non-response and other sampling and non-sampling errors. To correct for these biases, the CSO constructs a weighting mechanism to make the survey more representative of the true population. Prior to the third quarter (Q3) of 2006, the weights used by the CSO took no explicit account of immigration into Ireland. However, since Q3 2006, immigration has been factored into the *QNHS* weighting scheme.

The primary aim of this paper is to evaluate the effectiveness of both sets of the *QNHS* weighting mechanisms (pre-Q3-2006 and post-Q3-2006) for the study of immigrants in Ireland. Using published Census of Population figures as benchmarks, we compare them with corresponding weighted and unweighted estimates from the *QNHS* before and after Q3 2006, hence, we evaluate the effectiveness of both sets of the *QNHS* weighting schemes.

While this paper is principally concerned with examining the effectiveness of the *QNHS* weights for immigration research, we note that there has been no previously published evaluation of the weighting schemes utilised by the CSO, making this the first time the *QNHS* weighting schemes have been critically examined for their reliability in terms of matching the overall population.

Before discussing our research, we provide some background information on immigration in Ireland. Compared to nations of immigrants such as Canada, the United States and Australia, Ireland has only experienced large-scale immigration in the past two decades, notably following the enlargement of the European Union (EU) in 2004. Along with the United Kingdom (UK) and Sweden, Ireland was one of only three EU-15 countries that allowed nationals of the New Member States (NMS)[1] to freely access their labour markets. That dramatically changed Ireland's demographics as it welcomed an influx of immigrants from the accession countries. From a country with relatively few immigrants, Ireland has become one with a significant share of the immigrant population within the EU. By January 2011, 6.6 per cent of the EU-27's population were foreign nationals, whereas Ireland's foreign nationals made up 12 per cent of its population (Vasileva, 2012; Central Statistics Office, 2012).

Considering that foreign nationals compose a large portion of Ireland's population, the amount of research on the immigrant community is relatively modest and predominantly qualitative in nature. The lack of quantitative research on immigration in Ireland can be explained primarily by the absence

[1] New Member States (NMS) refer to EU15-25 countries between 2004 and 2006. From 2007 when Bulgaria and Romania joined the European Union, NMS refer to EU15-27 countries.

of rich data sources on immigrants. While countries like Canada and Australia have conducted major surveys targeting their immigrants such as the *Longitudinal Survey of Immigrants to Canada* and the *Longitudinal Survey of Immigrants to Australia*, there is not one national survey designed specifically to understand characteristics of the immigrant population in Ireland. Therefore, past quantitative research on immigration has to rely on data from general national surveys such as the Census of Population, the *EU-SILC*, the *NES* and the *QNHS*. Out of these, the *QNHS* is most widely used because it is conducted quarterly, has large sample sizes and contains important information on immigrants such as their demographics (i.e., nationality, place of birth, age, sex and marital status), labour market participation and education level.

Being a voluntary sample survey, the *QNHS* is prone to non-response and other sampling and non-sampling errors. To account for these, the CSO uses weights to match the *QNHS* results with known population estimates. When reviewing previous research on immigration in Ireland, we notice that the majority of past studies do not use the *QNHS* weights in their analysis. Specifically, while there have been a number of quantitative studies on characteristics of immigrants in Ireland such as Barrett *et al.* (2006), Duffy (2007), Barrett and Duffy (2008), Barrett and Kelly (2008), O'Connell and McGinnity (2008), Russell *et al.* (2008), Turner *et al.* (2008) and Barrett *et al.* (2011), Barrett and Kelly (2012), McGinnity *et al.* (2012), Kingston *et al.* (2013), only three of those studies (Russell *et al.*, 2008; Turner *et al.*, 2008; and McGinnity *et al.*, 2012) use the *QNHS* weights to obtain point estimates on Ireland's immigrant population.

It should be noted that despite the extensive use of the *QNHS* data, there has been only one study conducted by Barrett and Kelly (2008) to assess the reliability of the *QNHS* for immigration research. By comparing the profile of immigrants from the *Census of Population 2006* with the unweighted profile of immigrants from the *QNHS* 2006, the authors conclude that the *QNHS* is indeed a reliable source for migration research. In Section VII, we provide a comparison between our analysis and that of Barrett and Kelly (2008), as well as discuss the differences in results.

This paper is laid out as follows: starting from Sections I and II with background and introduction, Section III follows with details on the design of the *QNHS* and a thorough description of how the *QNHS* weights are constructed. In Sections IV and V, we will discuss how the pre- and post-Q3-2006 weights affect the reliability of the *QNHS* for immigration research. Finally, we will present our findings and discuss the use of the *QNHS* data for studying characteristics of immigrants in Ireland in Section VI and Section VII.

II INTRODUCTION AND DATA SOURCES

The main purpose of this paper is to explore the effectiveness of the *QNHS* weights for immigration research in Ireland. There are several possible definitions of "immigrants" so it is worth noting that we define immigrants as those who are not Irish in citizenship (i.e., foreign nationals). The reason for choosing this definition is to provide the best possible match with the post-Q3-2006 *QNHS* weighting scheme which takes nationality into account.

We start our analysis by comparing both the weighted and unweighted profile of immigrants from the second quarter (Q2) of the *QNHS* 2006 with the profile of immigrants from the *Census of Population 2006*, allowing us to assess the effectiveness of the pre-Q3-2006 *QNHS* weighting scheme. Since the Census is conducted in the second quarter (April) of the Census year, the results are comparable. We repeat the same analysis using the *QNHS* 2011 (Q2) and the *Census of Population 2011* to evaluate the post-Q3-2006 weighting methodology.

The *QNHS* 2006 (Q2) and the *QNHS* 2011 (Q2) samples were released to us by the Irish Social Science Data Archive (ISSDA),[2] and contain 85,314 and 59,361 observations respectively. The *Census 2006* and *Census 2011* figures are computed from the published tables on the CSO website.[3] There are a lot of variables available in the *QNHS* and the Census, however, the number of comparable variables between these two sources are very limited. For that reason, we only discuss the reliability of the *QNHS* for immigration research through five main variables of interest – age, sex, principal economic status, marital status and highest level of education attained.

III THE QUARTERLY NATIONAL HOUSEHOLD SURVEY

3.1 *Background and Sample Design*

The *QNHS* is a national household survey conducted quarterly by the CSO with the primary purpose of producing official labour estimates in Ireland. For their results to be generalisable, the *QNHS* samples have to be representative of the population with respect to characteristics of interest. Specifically, the CSO's goal is to obtain samples that mirror the population in terms of geographical region, age and sex composition. From Q3 2006, the CSO aims to obtain samples that are also representative of the population with regards to broad nationalities of residents.

[2] www.ucd.ie/issda
[3] http://www.cso.ie/en/census/index.html

The ideal way to achieve these representative samples would be to cross-stratify the population by geographical region, age and sex (and nationality), then draw independent samples from each stratum. However, without a population register that contains up-to-date demographics information (i.e., age, sex and nationality) on every member of a household, the CSO is unable to use this method. Therefore, in carrying out the *QNHS*, the CSO decides to stratify the country by geographical region and population density, and they count on large sample sizes and randomisation to achieve samples that are representative of the population regarding age and sex composition (and nationality).

The design of the *QNHS* is a two-stage stratified cluster sample design as described below (Central Statistics Office, 2011).[4]

- The sampling frame of households for the *QNHS* is obtained from the *Census of Population* and clustered into Primary Sampling Units (PSU) or geographical blocks, each containing an average of 75 dwellings.
- The sampling frame of households is also stratified into eight non-overlapping strata based on population density.
- In the first stage of sampling, a sample of 2,600 PSUs are selected at county level using Probability Proportional to Size Sampling. This means that the number of sampled PSUs in each stratum is proportional to the size of that stratum. For example, if 10 per cent of the total population come from Dublin region, 10 per cent of the total PSUs (or 260 PSUs) will be chosen from Dublin region.
- In the second stage of sampling, 15 households in each of the 2,600 PSUs are selected using Simple Random Sampling. That gives a total quarterly design sample of 39,000 households.
- Each household chosen for the *QNHS* is asked to participate for five consecutive quarters, with one-fifth of the total households replaced each quarter. That results in an overlap of 80 per cent in two consecutive quarters, and an overlap of 20 per cent between the same quarter in consecutive years.

The fact that a complete stratification is not used in the *QNHS* together with the fact that the survey suffers from non-response and random variation

[4] *Note:* Following *Census 2011*, a new *QNHS* sample design was introduced in Q4 2012. Details on this new design can be found at: http://www.cso.ie/en/releasesandpublications/er/*qnhs*/quarterlynationalhouseholdsurveyquarter12015. Since the *QNHS* data sets used in this paper are from 2006 and 2011 at which times the previous sample design was in use, we describe this previous sample design here.

means that the *QNHS* samples may differ from the population substantially. The CSO corrects for this discrepancy between the sample actually obtained and the ideal representative sample by weighting the actual sample to match the population in terms of variables of interest. We describe this weighting scheme in detail in the following sub-section.

3.2 *Weights in the QNHS*

With the design of the *QNHS*, every household in the sampling frame has an equal probability of selection. In an ideal world without non-response issues, every sampled household would have the same weight. For example, suppose we randomly survey 200 households from a population of 10,000 households, then every household in the population has an equal probability of 0.02 of being selected. If all 200 households respond, each household is given a weight of 10,000/200 or 50. It means that every sampled household represents 50 households in this hypothetical population. This is essentially the definition of household weight. It is a grossing factor and refers to the number of households in the population that each sampled household represents. Similarly, personal weight is interpreted as the number of people in the population that a sampled individual represents.

In reality, it is almost impossible for a survey to achieve 100 per cent response rate. Being a voluntary sample survey, the *QNHS* suffers from non-response and other sources of sampling and non-sampling errors. This leads to over-representation or under-representation of various subgroups. As a result, adjustment is performed on the *QNHS* sample obtained to ensure its representativeness. To do so, the CSO essentially assigns smaller weights to over-represented groups and larger weights to under-represented groups.

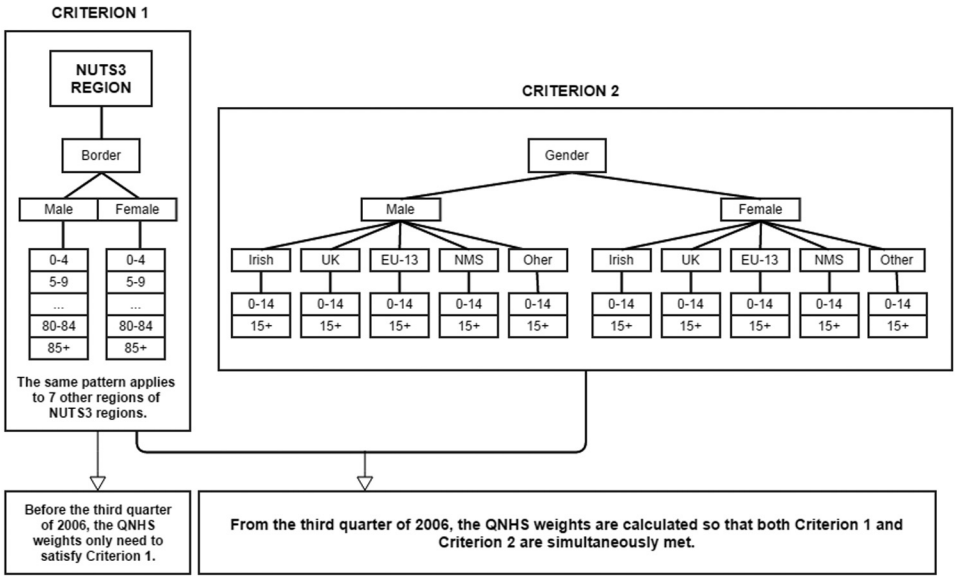
It is an intricate process to construct both household weights and personal weights for any complex survey design such as the *QNHS*. In this paper, we are only interested in evaluating the effectiveness of the *QNHS* personal weights for the study of immigrants in Ireland. While the household weights are an important component of the *QNHS*, they are not of interest to us at the moment. For that reason, the term "weights" in this paper always refer to personal weights.

The use of weights is entirely standard practice and survey statisticians often recommend using them to produce unbiased estimates of the population. By applying weights, the complexity of the sample design is considered, and non-response and other potential sources of errors are accounted for (Groves *et al.*, 2009, p. 328; Korn and Braubard, 1999, Section 4.6). However, when weighted estimates from a sample survey are similar to unweighted ones, using weights leads to higher sampling variance, hence, reduces the efficiency of the analysis (Korn and Graubard, 1999, Section 4.4).

Since its introduction in 1997, the *QNHS* has retained the same two-stage stratified cluster sample design, but its weighting scheme was modified in Q3 2006. Prior to Q3 2006, weights were calculated so that the *QNHS* results would agree with population estimates broken down by age (in five year increments from 0 to 85+ years), sex and NUTS3 region.[5] Since Q3 2006, the CSO have constructed weights using two different criteria.

The first criterion is the same as previously carried out: survey results are weighted to agree with population estimates by age, sex and NUTS3 region. The second criterion requires the *QNHS* results to match population estimates by age (under 15, 15 and above), sex and five broad nationality groups – Irish, UK, EU-13 (the original EU-15 excluding Ireland and the UK), NMS and Other Nationals. To ensure that the *QNHS* weights satisfy both criteria, the calibration process is performed using the CALMAR macro in SAS (Sautory, 1993). The construction of the *QNHS* weights is illustrated in Figure 1.

Figure 1: *Diagram of the Construction of the QNHS Weights*



The *QNHS* typically collects more than 100 variables covering a number of aspects. Out of all, the CSO chooses four variables – NUTS3 region, age, sex and nationality – to construct the *QNHS* weights as seen in Figure 1. We call these variables calibration variables because they explicitly form part of the

[5] The NUTS3 regions correspond to eight Regional Authorities: Border, Dublin, Midland, Mid-East, Mid-West, South-East, South-West and West.

weighting mechanism and are used to calibrate the *QNHS* samples to match population estimates. (Note that nationality is a calibration variable only from Q3 2006, but not before.) Apart from these four variables, all other variables are called non-calibration variables and do not contribute to the computation of the weights. In our paper, we investigate the reliability of the *QNHS* weights for immigration research through five main variables of interest. They include two calibration variables (age and sex) and three non-calibration variables (principal economic status, marital status and highest level of education attained).

Before conducting our analysis, it is worth summarising our expectations. The effectiveness of the *QNHS* weights for capturing the profile of immigrants in Ireland does not only depend on the weighting scheme (pre- or post-Q3-2006), but also whether the variable of interest is a calibration or non-calibration variable. We expect that the *QNHS* weights will perform well for calibration variables since the weights are specifically designed to create a match between the weighted *QNHS* sample and the Census for these variables. It is not clear, however, how the *QNHS* weights will perform on non-calibration variables and this is of course a question of interest to all researchers using the *QNHS* data. In Sections IV and V, we present our analysis on the effectiveness of both sets of the *QNHS* weighting mechanisms (pre- and post-Q3-2006) on calibration and non-calibration variables.

IV EFFECT OF THE PRE-Q3-2006 *QNHS* WEIGHTS ON IMMIGRATION RESEARCH

In this section, we compare weighted and unweighted profiles of immigrants from the Q2 2006 *QNHS* with the corresponding published figures from the *Census 2006* to evaluate the effectiveness of the pre-Q3-2006 *QNHS* weights. Recall that prior to Q3 2006, the *QNHS* weights are calculated so that the survey results will match population estimates broken down by age, sex and NUTS3 region. To start, we examine the nationality distribution in 2006 as shown in Table 1.

From Table 1, we see that the unweighted *QNHS* overcounts those with Irish nationality by 3.5 per cent but undercounts all other nationality groups. When weights are used, they help correct for the under-representation of these foreign national groups, resulting in a substantial decrease in sum of squared difference (SSE) between the *QNHS* estimates and the corresponding Census figures. Having a small SSE is always desirable because it indicates that the *QNHS* sample mirrors the Census closely. Here we see that the use of weights improves the *QNHS* estimated nationality distribution even though the pre-Q3-2006 weights do not explicitly take nationality into account.

Table 1: *Nationality Distribution in 2006 (%)*

Nationality	*Census*	*Unweighted QNHS*	*Weighted QNHS*
Irish	89.8	93.3	90.0
UK	2.7	1.9	2.5
EU-13	1.0	0.8	1.2
NMS	2.9	2.1	3.2
Other	3.6	2.0	3.1
SSE		16.1	0.5

Note: (Apply to all tables).
(1) Percentages have been rounded to the nearest decimal place and may not always amount to 100 per cent.
(2) SSE is the sum of squared differences between the Census figures and the corresponding weighted and unweighted *QNHS* estimates.

However, further analysis shows that, overall, the *QNHS* weights are often ineffective and lead to biased estimates of the main immigrant groups. We come to this conclusion by examining the effect of weights in twenty cases, which is the result of exploring the age, sex, principal economic status, marital status and highest level of education distributions of four foreign national groups (UK, EU-13, NMS and Other Nationals). In each case, we consider weights to be effective if there are less weighted *QNHS* estimates that are significantly different from the Census figures, or if the weighted results produce smaller SSEs than unweighted ones. Significance is determined using the modified Clopper-Pearson methodology (Korn and Graubard, 1998) which incorporates the complex survey design. We summarise our findings in Table 2.

From Table 2, we see that in nine out of twenty cases examined, using weights yields less reliable estimates than not using weights. In seven out of twenty cases, the weighted results are better than unweighted ones, and in four cases there is no meaningful difference between weighted and unweighted estimates. We conclude that the pre-Q3-2006 *QNHS* weights are generally inadequate in capturing the profile of immigrants in Ireland.

Even though it is necessary to determine the overall reliability of the pre-Q3-2006 *QNHS* weights, we obtain more interesting and informative results by dividing the variables of interest into two groups, calibration and non-calibration variables. As mentioned in Section 3.2, we expect that the *QNHS* weights would perform well on calibration variables, but we are unsure of their performance on non-calibration ones. Since our results for this section are found in Appendix A, we only discuss a sample of cases to illustrate our findings.

Table 2: *Summary of Findings on the Effect of the pre-Q3-2006 QNHS Weights on Four Main Immigrant Groups*

	UK	*EU-13*	*NMS*	*Other Nationals*
Calibration Variables				
Age	✓	✗	✗	✗
Sex	✗	✓	–	–
Non-Calibration Variables				
Principal Economic Status	✓	✗	–	–
Marital Status	✓	✗	✗	
Highest Education Level	✓	✗	✗	

Note:
✓ Weights lead to more reliable estimates.
✗ Weights lead to less reliable estimates.
– Virtually no difference between weighted and unweighted results.

4.1 *Effect of Pre-Q3-2006 Weights on Calibration Variables*

Out of our five main variables of interest, age and sex are two calibration variables because they are used in the construction of the pre-Q3-2006 *QNHS* weights. We now discuss the effect of these weights on each calibration variable.

Age Distribution: When we examine the age distribution of the four main immigrant groups, the unweighted *QNHS* estimates appear far more reliable with consistently smaller SSEs than the weighted ones for the EU-13, NMS and Other Nationals. In Table 3 we present the age distribution of the NMS nationals as an example.

From Table 3, we can see that the unweighted *QNHS* captures the age distribution of the NMS nationals very well, while the weighted *QNHS* significantly undercounts the 0-14 age group while significantly overcounting the 25-34 year age group. Moreover, the total SSE for the weighted *QNHS* is almost five times larger than that for the unweighted one. This strongly indicates that the pre-Q3-2006 weights are ineffective and should not be used to obtain the age distribution of the NMS nationals. Similar patterns are also observed for the EU-13 and Other nationals.

Sex Distribution: From Table 2, we see that there is no clear preference for the use of weights in obtaining the sex distribution of the four main immigrant groups. Out of all, only the UK nationals do not appear to favour the use of weights. We come to this conclusion by noticing that the weighted *QNHS* estimates for the sex distribution of the UK nationals yield a higher SSE than unweighted ones, as seen in Table 4.

Table 3: *Age Distribution of the NMS Nationals in 2006 (%)*

Age Group	*Census*	*Unweighted QNHS*	*Weighted QNHS*
0-14	8.3	7.7	6.6*
(s.e)		(0.8)	(0.7)
15-24	27.7	29.0	28.0
(s.e)		(1.3)	(1.3)
25-34	44.1	42.8	47.7*
(s.e)		(1.4)	(1.5)
35-49	16.5	16.7	14.5
(s.e)		(1.1)	(1.0)
50-64	3.1	3.8	3.1
(s.e)		(0.6)	(0.5)
65+	0.3	0.1	0.0*
(s.e)		(0.1)^{s}	(0.0)^{s}
SSE		4.3	20.0

Note: (Apply to all tables).
(1) The superscript *s* indicates cell sample size of less than 30.
(2) The *QNHS* estimate is significantly different from the corresponding Census figure at: {*} 0.05 significance level; {**} 0.01 significance level.

Table 4: *Sex Distribution of the UK Nationals in 2006 (%)*

Sex	*Census*	*Unweighted QNHS*	*Weighted QNHS*
Male	49.9	51.3	51.8
(s.e)		(1.1)	(1.1)
Female	50.1	48.7	48.2
(s.e)		(1.1)	(1.1)
SSE		3.9	7.0

So far we have discussed a sample of cases in which the pre-Q3-2006 *QNHS* weights are ineffective in capturing the age and sex distributions of the four foreign national groups. Refer back to Table 2, we see that out of eight cases (sex and age distributions of the four immigrant groups), four of them point to the ineffectiveness of the *QNHS* weights while only in two cases do weights improve the estimates. These two cases represent the age distribution of the UK and the sex distribution of the EU-13 nationals, whose estimates can be found in Table A1 and Table A2 in Appendix A.

4.2 *Effect of Pre-Q3-2006 Weights on Non-Calibration Variables*

When we examine the effect of the pre-Q3-2006 *QNHS* weights on three non-calibration variables – principal economic status, marital status and

highest level of education attained – an interesting picture emerges. A clear division of the four foreign national groups into two groups can be observed from Table 2. The first group consists of the UK and Other nationals, which favour the use of the *QNHS* weights. The other group includes the EU-13 and NMS nationals, which do not benefit from the *QNHS* weights. We now discuss the effect of the pre-Q3-2006 weights on the three non-calibration variables.

Principal Economic Status Distribution: When studying characteristics of immigrants, researchers are often interested in their performance and participation in the labour force. It should be noted that the CSO uses the International Labour Office (ILO) classification to calculate official employment and unemployment figures in Ireland. However, this information is not available in the Census, hence we examine the principal economic status of immigrants to understand their labour force participation.

We find that using weights makes no meaningful difference to the estimates of the principal economic status distributions of the NMS and Other Nationals. However, weights have contrasting effects on the EU-13 and UK groups. We show the principal economic status distributions of these two groups side-by-side in Table 5 for comparison.

Table 5: *Principal Economic Status Distribution of EU-13 and UK Nationals in 2006 (%)*

Economic Status	*Census*	*Unweighted QNHS*	*Weighted QNHS*	*Census*	*Unweighted QNHS*	*Weighted QNHS*
		EU-13			*UK*	
At Work	74.2	73.7	74.8	56.7	54.4	56.6
(s.e)		(2.0)	(2.0)		(1.4)	(1.4)
Unemployed	5.4	4.9	5.1	6.7	5.3	5.4
(s.e)		(0.9)	(1.0)		(0.6)	(0.7)
Student	8.5	11.9*	11.8*	6.9	6.6	6.3
(s.e)		(1.6)	(1.6)		(0.7)	(0.7)
Home Duties	6.8	6.7	6.0	13.0	17.0**	16.5**
(s.e)		(1.0)	(0.9)		(1.0)	(1.0)
Other	5.0	2.8*	2.3**	16.8	16.7	15.2
(s.e)		(0.7)	(0.6)		(1.1)	(1.0)
SSE		16.9	19.1		23.4	16.9

Note:
(1) These estimates are only for those who are 15 years and older.
(2) "Other" category refers to those who are retired, unable to work due to sickness/ illness, or other inactive people.

From Table 5, for the EU-13 nationals, we notice that both the weighted and unweighted *QNHS* significantly overcount the Student category while significantly undercounting those in Other category. However, the weighted *QNHS* yields a slightly higher SSE and a higher significance level for the Other category, we conclude that the unweighted results are preferred to the weighted ones for the EU-13 nationals. On the contrary, the UK nationals appear to benefit from the use of weights, as the weighted *QNHS* yields a smaller SSE than unweighted ones.

Marital Status Distribution: With regards to the marital status distribution, the pre-Q3-2006 *QNHS* weights again have a negative impact on the estimates of the EU-13 and NMS nationals, while a positive effect on the UK and Other Nationals. In Table 6, we show the marital status distributions of the EU-13 and Other Nationals to show the contrasting effects of the *QNHS* weights.

Table 6: *Marital Status Distribution of EU-13 and Other Nationals in 2006 (%)*

Marital Status	*Census*	*Unweighted QNHS*	*Weighted QNHS*	*Census*	*Unweighted QNHS*	*Weighted QNHS*
		EU-13			*Other Nationals*	
Single	65.7	67.5	70.3*	49.2	52.0*	51.1
(s.e)		(2.0)	(2.0)		(1.3)	(1.4)
Married	27.6	27.8	25.7	45.8	44.6	45.6
(s.e)		(1.9)	(1.9)		(1.3)	(1.4)
Divorced/ Separated	5.3	3.3*	2.9**	3.8	2.4**	2.3**
(s.e)		(0.7)	(0.6)		(0.4)	(0.4)
Widowed	1.4	1.4	1.1	1.2	1.0	1.0
(s.e)		(0.5)	(0.4)		(0.2)	(0.2)
SSE		7.3	30.6		4.1	5.9

We can see that for the EU-13 nationals, using weights does not only increase the total SSE by more than four times, but also leads to the Single category significantly different from the Census figure. Hence, we conclude that weights are ineffective in this case. However, the opposite is observed for the Other Nationals, as using weights helps correct for the significant over-representation of the Single category. Therefore, we favour the use of weights for the marital status distribution of the Other Nationals.

Highest Level of Education Distribution: When inspecting the highest level of education distribution of the main foreign national groups, we notice significant differences between the Census figures and the corresponding

QNHS ones (both weighted and unweighted), which shows the unreliability of the *QNHS* in capturing the highest level of education of the immigrants.

With regards to the effectiveness of the pre-Q3-2006 *QNHS* weights, we again find that the EU-13 and the NMS nationals do not benefit from using the *QNHS* weights, while the UK and Other Nationals do. In Table 7, we present the highest level of education distributions of the NMS and the UK nationals as an example. It shows that for the NMS group, applying weights leads to significantly unreliable estimates in all categories of the highest level of education attained, and a higher total SSE. It demonstrates the ineffectiveness of the *QNHS* weights for the NMS nationals. In contrast, for the UK nationals, we see that the weighted *QNHS* yields a much smaller total SSE than the unweighted *QNHS*, hence weights are preferred in this case.

Table 7: *Highest Education Attainment Distribution of NMS and UK Nationals in 2006 (%)*

Education	*Census*	*Unweighted QNHS*	*Weighted QNHS*	*Census*	*Unweighted QNHS*	*Weighted QNHS*
		NMS			*UK*	
Primary	4.6	9.8**	9.0**	8.0	16.0**	14.8**
(s.e)		(1.2)	(1.2)		(1.1)	(1.1)
All Secondary	69.7	61.4**	60.7**	54.8	48.8**	48.6**
(s.e)		(1.9)	(1.9)		(1.5)	(1.5)
All Third Levels	25.7	28.8	30.3*	37.2	35.2	36.7
(s.e)		(1.7)	(1.8)		(1.4)	(1.5)
SSE		105.5	121.5		104.0	84.9

Note: These estimates are for those aged 15 years and over, and whose education has ceased.

4.3 *Remarks*

In this section, we have investigated the effectiveness of the pre-Q3-2006 *QNHS* weights for immigration research in Ireland. After examining the age, sex, principal economic status, marital status and highest level of education distributions of the four main immigrant groups, we find that the pre-Q3-2006 weights are overall ineffective and unreliable.

When we divide the five variables of interest into two groups, with the first group consisting of calibration variables (age and sex) and the second group consisting of non-calibration variables (principal economic status, marital status and highest level of education), we observe some interesting results.

For calibration variables, one would expect that the pre-Q3-2006 *QNHS* weights would perform well because both age and sex variables are used to

construct these weights. However, this is not the case. After examining the age and sex distributions of the four immigrant groups, we find that half of them point to the ineffectiveness of the pre-Q3-2006 weights, while only a quarter of them show improvement in estimates when weights are used. This may seem counter intuitive, but an explanation for this result is that the nationality variable was not included in the construction of the pre-Q3-2006 weights. Hence, when the response rate of each immigrant group broken down by age and sex is different from the corresponding response rate of the population, grossing up the *QNHS* results to match population estimates would bias estimates of that immigrant group. This leads to the pre-Q3-2006 *QNHS* weights being ineffective in half of the cases examined. When the disparity between the response rate distribution of immigrants and that of the population is small, weights would perform better, as seen in the minority of cases.

For non-calibration variables, we notice an unexpected pattern in the effect of the pre-Q3-2006 *QNHS* weights on the four immigrant groups. When we examine the principal economic status, marital status and highest level of education distributions of the four foreign national groups, we see a clear division into two groups. The first group includes the UK and Other Nationals, who appear to benefit from the use of the pre-Q3-2006 *QNHS* weights consistently. Contrastingly, the *QNHS* weights have consistently negative effects on all three non-calibration variable distributions of the EU-13 and NMS nationals. This clear distinction indicates certain similarity between the UK and Other Nationals, and between the EU-13 and NMS nationals.

Overall, the pre-Q3-2006 *QNHS* weights are not very reliable when used to obtain estimates on the immigrant population in Ireland. Even in cases where weights do improve these estimates, the improvement is very minor. When conducting the analysis in this section, we also observe that there are many weighted and unweighted estimates that are significantly different from the corresponding Census figures. It means that both the weighted and unweighted *QNHS* are unreliable for immigration research. This is a concern because the *QNHS* weights are not doing what they are supposed to do, which is to make the sample more representative of the true population. Therefore, certain analysis performed on either weighted or unweighted data will not be generalisable.

The CSO also recognised that the *QNHS* did not adequately capture the immigrant population or the demographic shift in Ireland. As a result, in Q3 2006, they modified the calculation of the *QNHS* weights to take into account nationalities of respondents. In the next subsection we consider how effectively these new weights perform.

V EFFECT OF THE POST-Q3-2006 *QNHS* WEIGHTS ON IMMIGRATION RESEARCH

From 2006 Q3, the *QNHS* weights are constructed with two criteria that the weighted *QNHS* results have to match population estimates broken down by age, sex and region composition, as well as by age (below 15, 15 and above), sex, and five broad nationality groups (Irish, UK, EU-13, NMS and Other Nationals). In this section, we evaluate the revised *QNHS* weights by replicating the same analysis from Section IV using the published *Census 2011* figures and the *QNHS* 2011 (Q2) sample.

We start our analysis by examining the nationality distribution in 2011. Table 8 shows that the unweighted *QNHS* overcounts the Irish nationality group and undercounts all other immigrant groups. When weights are applied, there is absolutely no difference between the Census and the corresponding weighted *QNHS* figures, and the *QNHS* is now representative of the population of Ireland with regards to nationality distribution.

This is not surprising because the *QNHS* weights are constructed to match the *QNHS* results with population estimates broken down by these five nationality groups.

Table 8: *Nationality Distribution in 2011 (%)*

Nationality	*Census*	*Unweighted QNHS*	*Weighted QNHS*
Ireland	87.8	90.7	87.8
UK	2.5	1.8	2.5
EU-13	1.1	0.7	1.1
NMS	5.1	4.2	5.1
Other	3.5	2.6	3.5
SSE		10.7	0.0

After examining the sex, age, principal economic status, marital status and highest level of education distributions of the four main immigrant groups, overall, we find an improvement in the performance of the post-Q3-2006 *QNHS* weights. Out of 20 cases examined, weights produce better estimates of the immigrant groups in 11 cases. In five cases, weights lead to worse estimates of the foreign national groups, and in four cases there is virtually no difference between weighted and unweighted results. We summarise our findings in Table 9.

Table 9: *Summary of Findings on Whether Weights Produce More Reliable Estimates of the Immigrant Population in Ireland 2011*

	UK	*EU-13*	*NMS*	*Other Nationals*
Calibration Variables				
Age	✓	✓	✓	✓
Sex	–	✓	–	✓
Non-Calibration Variables				
Principal Economic Status	✗	✓	✓	✓
Marital Status	✗	✗	–	✗
Education	✗	✓	–	✓

Note:
✓ Weights lead to more reliable estimates.
✗ Weights lead to less reliable estimates.
– There is virtually no difference between weighted and unweighted results.

Even though the overall performance of the post-Q3-2006 *QNHS* weights has improved, we gain more insight behind this performance by looking at effects of these modified *QNHS* weights on calibration and non-calibration variables separately. As all of our results from this section are shown in Appendix B, we again only discuss a sample of cases here to illustrate our findings.

5.1 *Effect of Post-Q3-2006 Weights on Calibration Variables*

Age Distribution: To begin, we notice that applying weights generates more reliable estimates of the age distribution of all four immigrant groups, with consistently smaller SSEs than corresponding unweighted ones. We show the age distribution of the UK nationals in Table 10 as an example. It shows that the unweighted *QNHS* significantly overcounts the 0-14 group while undercounting the 25-34 age group, but this is corrected when weights are used. Similar patterns are also observed for other immigrant groups.

Sex Distribution: Our findings show that the weighted *QNHS* captures the sex distribution of the foreign national groups well, with consistently smaller SSEs or at most equal to those for unweighted estimates. For instance, in Table 11 we present the sex distribution of the EU-13 nationals. The unweighted *QNHS* significantly miscount the sex distribution of the EU-13 nationals, but the weighted *QNHS* correct for this resulting in negligible SSE.

As seen in Table 9, other than the two cases in which there is no meaningful difference between the weighted and unweighted estimates for age and sex distributions of the immigrant groups, the post-Q3-2006 *QNHS* weights

improve the estimates in the remaining six out of eight cases. This indicates that the post-Q3-2006 weights are reliable and should be used to obtain the distributions of calibration variables of the immigrant groups.

Table 10: *Age Distribution of the UK National in 2011 (%)*

Age Group	*Census*	*Unweighted QNHS*	*Weighted QNHS*
0-14	9.9	12.8*	9.9
(s.e)		(1.4)	(1.1)
15-24	10.6	9.9	10.3
(s.e)		(1.0)	(1.1)
25-34	12.2	9.2**	10.8
(s.e)		(0.9)	(1.1)
35-49	33.1	33.2	33.8
(s.e)		(1.5)	(1.5)
50-64	21.5	20.8	20.9
(s.e)		(1.4)	(1.4)
65+	12.7	14.1	14.3
(s.e)		(1.2)	(1.3)
SSE		20.4	5.5

Table 11: *Sex Distribution of the EU-13 Nationals in 2011 (%)*

Sex	*Census*	*Unweighted QNHS*	*Weighted QNHS*
Male	48.3	43.4*	48.4
(s.e)		(2.2)	(2.3)
Female	51.7	56.6*	51.6
(s.e)		(2.2)	(2.3)
SSE		48.0	<0.1

5.2 *Effect of Post-Q3-2006 Weights on Non-Calibration Variables*

In the previous subsection, we conclude that the post-Q3-2006 *QNHS* weights have a positive effect on both calibration variables – age and sex, which seemingly indicates an overall improvement in construction of the *QNHS* weights. However, this is not the case when we examine their effect on non-calibration variables.

Referring back to Table 9, we observe that when examining the principal economic status, marital status and highest level of education attained for the four immigrant groups, weights only improve the estimates in half of cases and

worsen estimates in the other half. There is no improvement in the performance of the post-Q3-2006 *QNHS* weights on non-calibration variables.

An interesting thing to observe from Table 9 is that while the performance of the post-Q3-2006 weights on the EU-13, NMS and Other Nationals depends on the non-calibration variable itself, the UK nationals appear to suffer from the use of weights consistently. We now examine the effect of the post-Q3-2006 *QNHS* weights for each non-calibration in detail.

Principal Economic Status Distribution: When examining the principal economic status of the immigrant groups, we notice a number of things. First of all, weights do not substantially improve the reliability of the *QNHS* estimates. In most cases, when the unweighted *QNHS* estimate is found to deviate from the Census significantly, so does the corresponding weighted estimate. However, weighted estimates still yield smaller SSEs for three out of four immigrant groups (except the UK).

The second thing we notice when exploring the economic status distribution for all immigrant groups is that the *QNHS*, whether weighted or unweighted, significantly overcounts the Home Duties category for all immigrant groups. We then compute the economic status distribution of the Irish nationals and of the entire *QNHS* sample and observe that the Home Duties category is also significantly overcounted in both cases. We conclude that this is rather a problem with the voluntary nature of the *QNHS* than a problem with any particular nationality group. To demonstrate, we present the principal economic status distribution of the Other Nationals in Table 12.

Table 12: *Principal Economic Status Distribution of the Other Nationals in 2011 (%)*

Economic Status	*Census*	*Unweighted QNHS*	*Weighted QNHS*
At Work	48.9	52.4*	52.4*
(s.e)		(1.6)	(1.6)
Unemployed	16.5	9.1**	9.3**
(s.e)		(0.8)	(0.9)
Student	20.4	18.0	18.1
(s.e)		(1.3)	(1.4)
Home Duties	9.3	16.9**	16.5**
(s.e)		(1.0)	(1.0)
Other	5.0	3.6*	3.6*
(s.e)		(0.5)	(0.5)
SSE		132.5	123.2

Note: These estimates are only for those who are 15 years and older.

Due to a smaller total SSE seen in Table 12, we conclude that the weighted estimates are preferred to the unweighted ones. However, we note that our conclusion does not imply that the weighted estimates are reliable, as Table 12 shows that the weighted *QNHS* still fails to capture every category but one (i.e., Student). This result is rather discouraging especially to social scientists who want to study the economic status of this minority group. In this case, weights do not help make the *QNHS* sample any more representative of the Other Nationals population, hence, any analysis on this group would not be generalisable.

Marital Status Distribution: Except for the NMS nationals where there is no meaningful difference between the weighted and unweighted estimates, the post-Q3-2006 *QNHS* weights do not benefit any of the other three immigrant groups. We show the marital status distribution of the UK nationals in Table 13 for illustration. It can be seen that the unweighted *QNHS* captures the marital status distribution of the UK nationals very well, while the weighted *QNHS* significantly overcounts the Married category, and produces a higher SSE. Similar patterns are observed for the EU-13 and Other Nationals.

Highest Level of Education Distribution: When inspecting the highest level of education distributions of the four main immigrant groups, we see that the post-Q3-2006 *QNHS* weights improve estimates for the EU-13 and Other Nationals, while those for the UK nationals worsen. In Table 14, we show the highest level of education distribution for the EU-13 and UK nationals side-by-side to show the contrasting effects of the post-Q3-2006 *QNHS* weights. For the EU-13 nationals, we see that weighted *QNHS* results in a smaller SSE, while the opposite is seen for the UK nationals.

Table 13: *Marital Status Distribution of the UK Nationals in 2011 (%)*

Economic Status	*Census*	*Unweighted QNHS*	*Weighted QNHS*
Single	39.4	39.1	37.8
(s.e)		(1.7)	(1.7)
Married	46.6	49.3	50.6*
(s.e)		(1.7)	(1.7)
Divorced/Separated	9.9	8.3	8.3
(s.e)		(0.9)	(0.9)
Widowed	4.2	3.3	3.3
(s.e)		(0.6)	(0.6)
SSE		10.8	21.9

Table 14: *Highest Education Level Distribution of the EU-13 and UK Nationals in 2011 (%)*

Education Level	*Census*	*Unweighted QNHS*	*Weighted QNHS*	*Census*	*Unweighted QNHS*	*Weighted QNHS*
		EU-13			*UK*	
Primary	3.3	3.8	3.5	9.7	9.4	9.1
(s.e)		(1.0)[s]	(1.0)[s]		(1.0)	(1.0)
All Secondary	35.6	37.0	35.2	56.1	50.1*	49.7*
(s.e)		(2.7)	(2.8)		(1.8)	(1.9)
All Third Levels	61.1	59.2	61.3	34.2	40.6*	41.2*
(s.e)		(2.9)	(3.0)		(1.8)	(1.8)
SSE		5.8	0.2		77.1	90.3

Note: These estimates are only for those aged 15 years and older whose education has ceased.

5.3 *Remarks*

In this section we have examined the effects of the post-Q3-2006 *QNHS* weights for immigration research in Ireland. Overall, we find that there is an improvement in the performance of the post-Q3-2006 weights, compared to the pre-Q3-2006 ones. After exploring the five main variable distributions of the four immigrant groups, we see that the post-Q3-2006 weights improve estimates in 11 out of 20 cases, while estimates in five cases worsen. To understand the performance of the post-Q3-2006 weights better, we divide the five variables of interest into calibration and non-calibration variable groups, and observe a number of things.

For calibration variables, the post-Q3-2006 *QNHS* weights are effective in capturing the age and sex distributions of all four immigrant groups. This is of course not surprising because from Q3 2006, the *QNHS* weights are especially calculated to match population estimates broken down by age, sex and main nationality groups. We conclude that post-Q3-2006 *QNHS* weights are reliable for calibration variables, and should be used when researchers are interested in exploring the age and sex distributions of the immigrant population in Ireland.

For non-calibration variables, we see no improvement in the performance of the post-Q3-2006 *QNHS* weights. In 12 cases examined (as a result of examining the principal economic status, marital status and highest level of education distributions of the four main immigrant groups), weights lead to more reliable estimates in five cases while they worsen estimates in another five cases. It demonstrates that the post-2006-*QNHS* weights are still not reliable, and it remains unclear whether one should use weights to obtain the distributions of non-calibration variables for immigration research.

One interesting thing that we observe when investing the effects of the post-Q3-2006 *QNHS* weights on non-calibration variables is the difference between the UK nationals and the other three immigrant groups. We find that the UK nationals do not benefit from the use of post-Q3-2006 weights consistently, while there is no similar pattern observed for the other three groups. We provide a hypothesis for why the UK nationals stand out among the immigrant groups in Section VI.

VI CONCLUSIONS

Large-scale immigration is a relatively new phenomenon in Ireland. After the EU enlargement in 2004, an influx of immigrants from accession countries came to Ireland due to its open labour market policy. Since then, there have been a number of quantitative studies on characteristics of the immigrant population in Ireland, and they have used the *QNHS* as the main data source.

In this paper, we have investigated the reliability of both sets of the *QNHS* weighting mechanisms (pre-Q3-2006 and post-Q3-2006) for immigration research. Since the *QNHS* was introduced in 1997, the construction of the *QNHS* weights has been modified once in Q3 2006. Before Q3 2006, the weights were calculated so that the results would match population estimates broken down by age, sex and region. From Q3 2006, the weights have been constructed to not only match population estimates broken down by age, sex and region, but also to match population estimates broken down by age, sex and broad nationality groups. The nationality variable has been included in the construction of the post-Q3-2006 *QNHS* weights to take into account the demographic shift in Ireland's population post-EU enlargement.

Our analysis shows that the effectiveness of the *QNHS* weights does not only depend on the weighting scheme used (pre- versus post-Q3 2006) but also depends on whether the variable is calibration or non-calibration variable. We examine five variables of interest throughout the paper. Of these variables, age and sex are two calibration variables, while principal economic status, marital status and highest level of education attained are three non-calibration variables.

For calibration variables, we find that the pre-Q3-2006 weights are not adequate in capturing the age and sex distributions of the four main immigrant groups, namely the UK, EU-13, NMS and Other Nationals. Our explanation for this result is that the nationality variable was not included in the construction of the pre-Q3-2006 weights. Therefore, when the response rate of each immigrant group broken down by age and sex is different from the corresponding response rate of the population, grossing up the *QNHS* results

to match population estimates introduces bias in estimates of that immigrant group. This leads to the pre-Q3-2006 *QNHS* weights being ineffective, and should not be used.

As a result, with the nationality variable included in the construction of the post-Q3-2006 weights, we expected that the post-Q3-2006 *QNHS* weights would be very effective in capturing the age and sex distribution of the main immigrant groups. The result is exactly what we expected. Therefore, we recommend the use of the post-Q3-2006 weights to obtain the age and sex distributions, or the calibration variable distributions, of the immigrant population in Ireland.

For non-calibration variables, it was not known how well the *QNHS* weights would perform. Ideally, properly constructed *QNHS* weights would make the samples more representative of the population with respect to both calibration and non-calibration variables. However, our analysis shows that both the pre-Q3-2006 and the post-Q3-2006 weights are unreliable when used to obtain various non-calibration distributions of the immigrant groups. There is no overall improvement in the performance of the *QNHS* weights after the modification in Q3 2006.

A few interesting results are observed as we examine the principal economic status, the marital status and the highest level of education distributions of the four foreign national groups. First of all, there is a clear division of the four immigrant groups into two groups regarding how they respond to the use of the pre-Q3-2006 *QNHS* weights. While the UK and Other Nationals benefit from using the pre-Q3-2006 *QNHS* weights, the EU-13 and NMS nationals do not. It signifies some similarity between the UK and Other Nationals, and between the EU-13 and NMS nationals.

Second of all, when the same analysis is conducted using the post-Q3-2006 weights, another pattern is observed. There is still a division of the four immigrant groups into two, but this time the UK nationals stand by themselves while the remaining group consists of the EU-13, NMS and Other Nationals. While the effect of the post-Q3-2006 weights on the latter group depends on the non-calibration variable being examined, the use of the post-Q3-2006 *QNHS* weights does not benefit the UK nationals consistently. It is in complete opposition to how the UK nationals respond positively to the pre-Q3-2006 weights. This is puzzling to us, as it means when the nationality variable is included in the construction of the post-Q3-2006 weights, somehow it has an opposite (and negative) effect on the UK nationals.

Throughout our analysis, we have seen that the UK nationals often stand out when compared with other immigrant groups. This is possibly due to the similarity between the Irish nationals and the UK nationals. Since the pre-Q3-2006 weighting scheme is entirely designed to match the population which is

made up of mostly Irish nationals, an immigrant group having similar characteristics to the Irish nationals such as the UK would benefit from using the pre-Q3-2006 weights.

This certainly does not explain why the post-Q3-2006 weights do not work as well for the UK nationals. What we know is that the difference between the pre-Q3-2006 and the post-Q3-2006 weighting scheme is the introduction of "nationality" as a calibration variable. Perhaps the issue here lies with the definition of "immigrants" itself. Recall that we define immigrants by their nationalities, so the UK nationals refer to those who define themselves as having UK citizenship. While this does not appear to be an issue with other groups, it could potentially be an issue with the UK nationals.

In our *QNHS* 2011 (Q2) sample, there are in total 1,870 people who define themselves as Irish nationals but were born in the UK. This is twice the sample size of those who are UK nationals and were born in the UK (933 people). We would expect many of those 1,870 people to have dual citizenships, so if there is a discrepancy in how one describes oneself in the Census and in the *QNHS* (i.e., one describes oneself as Irish in the Census and as British in the *QNHS*), it will cause the discrepancy between the Census and the *QNHS* estimates. Since the *QNHS* questionnaire does not provide an option for survey participants to claim dual citizenships, it is not possible for us at the moment to test this hypothesis. However, it would be definitely worth re-examining the effects of the post-Q3-2006 weights on the UK nationals, either with the *Census 2016* data when they come out or with simulation exercises.

VII DISCUSSION

We noted in Section 1 that while the *QNHS* data have been widely used, there has only been one analysis in the literature to date (Barrett and Kelly, 2008) to assess the reliability of the *QNHS* data for immigration research in Ireland. We now compare the differences in our analysis and theirs.

First, Barrett and Kelly (2008) restrict their analysis to labour force participants (those who are aged 15 to 64 and describe themselves as either employed or unemployed but looking for work). We do not impose this restriction on our data sets because we are interested in the overall reliability of the *QNHS* for research on the entire population of immigrants not just those in the labour force.

Second, while Barrett and Kelly (2008) define immigrants as those who are not Irish in citizenship and born outside of Ireland, we define immigrants as foreign nationals (i.e., those who are not Irish in citizenship). Their paper uses data from Q2 2006 and with this data there is no clear reason why either

definition is preferred. However, since nationality was incorporated in the weighting scheme in Q3 2006, it is important to analyse the post-Q3-2006 data using nationality as the defining characteristic of immigrants. We use this definition throughout our analysis to enable direct comparisons to be made. We also find that our choice to define immigrants by their nationality allows a direct comparison with the Census data.

Third, Barrett and Kelly (2008) use results from the 5 per cent Anonymised Records of *Census 2006* to compare with those from the *QNHS* 2006 (Q2). This sample drawn from the Census is of course subject to random sampling variation which impacts on its role as a gold standard against which the *QNHS* should be compared. Our gold standard is published Census results which are computed using the entire *Census 2006* data set. This allows us to provide more accurate benchmark figures because there is no random variation as a result of taking a 5 per cent sample of the population.

Finally, in their paper, Barrett and Kelly do not utilise weights for analysis but they indicate that weighted data would produce similar results to unweighted data. Even though our findings show that there are differences between weighted and unweighted analysis, it does not necessarily contradict the work of Barrett and Kelly (2008) considering the mentioned differences in our set up.

Our analysis shows that the (unweighted) *QNHS* is far from perfect in capturing the profile of immigrants in Ireland. There are many cases in which the unweighted *QNHS* estimates significantly depart from the corresponding Census figures. The CSO certainly does not suggest that the unweighted *QNHS* is representative of the population. There are several reasons why this is the case, first, the survey design only stratifies on geographical location and population density, and not on other important demographic variables. Second, the voluntary nature of the *QNHS* makes it prone to non-response and subsequently leads to the unweighted *QNHS* samples being unrepresentative of the population. To correct for these sources of bias and other sampling and non-sampling errors, the CSO construct the *QNHS* weights with the goal that these weights, when used, will make the *QNHS* samples more representative of the true population.

This paper has detailed how the pre-Q3-2006 and the post-Q3-2006 *QNHS* weights are constructed, and their effects when used to study characteristics of immigrants in Ireland. We conclude that the pre-Q3-2006 weights are not reliable for this purpose. There is an overall improvement in the performance of the post-Q3-2006 *QNHS* weights after the introduction of the "nationality" variable in the weighting scheme. However, there are still a number of cases in which the weighted estimates, though slightly better than the unweighted estimates, fail to correct for the bias of the *QNHS*.

It is our belief that the CSO should revisit both the stratification scheme and the weighting mechanism used for the *QNHS*. At the time of writing, the CSO is in the process of changing its stratification scheme. Instead of using geographical region and population density as stratification variables, they will use geographical region and the Haase-Pratschke Index of Relative Affluence and Deprivation (Haase and Pratschke, 2012) to stratify the country. As we expect the Haase-Pratschke *Index of Relative Affluence and Deprivation* to be available from *Census 2016*, a new *QNHS* design will hopefully be in place soon afterwards. It will definitely be interesting to see how this change improves the estimates of the *QNHS* for immigration research in Ireland.

In conclusion, we recommend researchers in the field to use the post-Q3-2006 weights in their analysis of the immigrant population in Ireland, but to be aware of the potential effects it can have on various outcome variables.

REFERENCES

BARRETT, A. and D. DUFFY, 2008. "Are Ireland's Immigrants Integrating into its Labor Market?", *International Migration Review*, Vol. 42, No. 3, pp. 597-619. doi: 10.1111/j.1747-7379.2008.00139.x

BARRETT, A. and E. KELLY, 2008. "How Reliable is the *Quarterly National Household Survey* for Migration Research?", *The Economic and Social Review*, Vol. 39, No. 3, pp. 191-205.

BARRETT, A. and E. KELLY, 2012. "The Impact of Ireland's Recession on the Labour Market Outcomes of its Immigrants", *European Journal of Population*, Vol 28, No 1, 2012, pp. 99-111. http://dx.doi.org/10.1007/s10680-011-9249-7

BARRETT, A., A. BERGIN and D. DUFFY, 2006. "The Labour Market Characteristics and Labour Market Impacts of Immigrants in Ireland", *The Economic and Social Review*, Vol. 37, No. 1, pp. 1-26.

BARRETT, A., A. BERGIN and E. KELLY, 2011. "Estimate the Impact of Immigration on Wages in Ireland", *The Economic and Social Review*, Vol. 42, No. 1, pp. 1-26.

CENTRAL STATISTICS OFFICE, 2011. "*Quarterly National Household Survey Quarter 2 2011*", Dublin. http://www.cso.ie/en/media/csoie/ releases publications/documents/ labourmarket/2011/*qnhs*_q22011.pdf (Accessed 19 August 2014).

CENTRAL STATISTICS OFFICE, 2012. *This is Ireland, Highlights from Census 2011, Part 1*, Dublin: Stationery Office.

DUFFY, D., 2007. "The Housing Tenure of Immigrants in Ireland: Some Preliminary Analysis", Working Paper No. 188, Dublin: The Economic and Social Research Institute (ESRI).

GROVES, R., F. FOWLER, M. COUPER, J. LEPKOWSKI, E. SINGER and R. TOURANGEAU, 2009. *Survey Methodology*, Second Edition, John Wiley & Sons.

HAASE, T. and J. PRATSCHKE, 2012. *Optimising the Sampling Methodology for CSO Household Surveys Theoretical Considerations and Simulation Study*, Dublin: Central Statistics Office.

KINGSTON, G., P. J. O'CONNELL and E. KELLY, 2013. *Ethnicity and Nationality in the Irish Labour Market: Evidence from the QNHS Equality Module 2010*, Dublin: Equality Authority/The Economic and Social Research Institute.

KORN, E. L. and B. I. GRAUBARD, 1998. "Confidence Intervals for Proportions With Small Expected Number of Positive Counts Estimated From Survey Data," *Survey Methodology*, Vol. 24, pp. 193-201.

KORN, E. and B. GRAUBARD, 1999. *Analysis of Health Surveys*, John Wiley & Sons.

MCGINNITY, F., D. WATSON and G. KINGSTON, 2012. *Analysing the Experience of Discrimination in Ireland: Evidence from the QNHS Equality Module 2010*, Dublin: Equality Authority/The Economic and Social Research Institute.

O'CONNELL, P. J. and F. MCGINNITY, 2008. *Immigrants at Work: Nationality and Ethnicity in the Irish Labour Market*, Dublin: Equality Authority/The Economic and Social Research Institute.

RUSSELL, H., E. QUINN, R. KING-O'RIAIN and F. MCGINNITY, 2008. *The Experience of Discrimination in Ireland: Analysis of the QNHS Equality Module*, Dublin: The Equality Authority and The Economic and Social Research Institute.

SAUTORY, O., 1993. "La macro CALMAR. Redressement d'un échantillon par calage sur les marges", Institut national de la satistique et des études économiques. http://www.insee.fr/fr/methodes/outils/calmar/doccalmar.pdf (Accessed 24 October 2014).

TURNER, T., M. D'ART and M. O'SULLIVAN, 2008. "Union Availability, Union Membership and Immigrant Workers", *Employee Relations*, Vol. 30, No. 5, pp. 479- 493.

VASILEVA, K., 2012. "Nearly Two-Thirds of the Foreigners Living in EU Member States are Citizens of Countries Outside the EU-27", *Population and Social Conditions*, EuroStat Statistics in focus.

APPENDIX A

2006 ESTIMATES

Table A1: *Sex Distribution of Main Immigrant Groups in 2006 (%)*

Sex	*Census*	*Unweighted QNHS*	*Weighted QNHS*	*Census*	*Unweighted QNHS*	*Weighted QNHS*
		UK			*EU-13*	
Male	49.9	51.3	51.8	48.6	44.0*	44.5*
(s.e)		(1.1)	(1.1)		(1.9)	(1.9)
Female	50.1	48.7	48.2	51.4	56.0*	55.5*
(s.e)		(1.1)	(1.1)		(1.9)	(1.9)
SSE		3.9	7.0		42.3	33.6
		NMS			*Other Nationals*	
Male	60.6	62.2	62.2	51.2	50.5	50.7
(s.e)		(1.1)	(1.1)		(1.2)	(1.2)
Female	39.4	37.8	37.8	48.8	49.5	49.3
(s.e)		(1.1)	(1.1)		(1.2)	(1.2)
SSE		5.1	5.1		1.0	0.5

Note: The Census figures are extracted from Table C0436 (*Census 2006* Published Tables).

Table A2: *Age Distribution of Main Immigrant Groups in 2006 (%)*

Age Group	*Census*	*Unweighted QNHS*	*Weighted QNHS*	*Census*	*Unweighted QNHS*	*Weighted QNHS*
		UK			*EU-13*	
0-14	13.9	16.2*	14.9	7.1	7.6	6.5
(s.e)		(1.1)	(1.1)		(1.4)	(1.2)
15-24	10.1	7.8**	7.9**	16.6	18.6	17.7
(s.e)		(0.7)	(0.7)		(1.7)	(1.6)
25-34	15.4	13.1*	16.9	40.7	37.0	43.3
(s.e)		(0.9)	(1.1)		(2.2)	(2.3)
35-49	32.3	31.1	31.1	23.7	24.6	22.9
(s.e)		(1.1)	(1.2)		(1.7)	(1.7)
50-64	18.6	20.2	18.6	8.5	9.5	7.6
(s.e)		(1.1)	(1.1)		(1.3)	(1.1)
65+	9.6	11.6*	10.6	3.4	2.7	2.1*
(s.e)		(0.9)	(0.8)		(0.7)	(0.5)
SSE		23.9	10.1		11.8	11.5
		NMS			*Other Nationals*	
0-14	8.3	7.7	6.6*	16.5	18.3	15.9
(s.e)		(0.8)	(0.7)		(1.1)	(1.0)
15-24	27.7	29.0	28.0	16.6	16.4	16.1
(s.e)		(1.3)	(1.3)		(1.1)	(1.2)
25-34	44.1	42.8	47.7*	36.1	34.4	39.7*
(s.e)		(1.4)	(1.5)		(1.4)	(1.5)
35-49	16.5	16.7	14.5	24.7	26.0	24.3
(s.e)		(1.1)	(1.0)		(1.1)	(1.1)
50-64	3.1	3.8	3.1	4.3	3.5	2.9*
(s.e)		(0.6)	(0.5)		(0.5)	(0.4)
65+	0.3	0.1	0.0*	1.8	1.3	1.1*
(s.e)		(0.1)	(0.0)		(0.3)	(0.3)
SSE		4.3	20.0		8.8	16.1

Note: The Census figures are extracted from Table C0436 (*Census 2006* Published Tables).

Table A3: *Principal Economic Status Distribution of Immigrant Groups 2006 (%)*

Economic Status	*Census*	*Unweighted QNHS*	*Weighted QNHS*	*Census*	*Unweighted QNHS*	*Weighted QNHS*
		UK			*EU-13*	
At Work	56.7	54.4	56.6	74.2	73.7	74.8
(s.e)		(1.4)	(1.4)		(2.0)	(2.0)
Unemployed	6.7	5.3	5.4	5.4	4.9	5.1
(s.e)		(0.6)	(0.7)		(0.9)	(1.0)
Student	6.9	6.6	6.3	8.5	11.9*	11.8*
(s.e)		(0.7)	(0.7)		(1.6)	(1.6)
Home Duties	13.0	17.0**	16.5**	6.8	6.7	6.0
(s.e)		(1.0)	(1.0)		(1.0)	(0.9)
Other	16.8	16.7	15.2	5.0	2.8*	2.3**
(s.e)		(1.1)	(1.0)		(0.7)	(0.6)
SSE		23.4	16.9		16.9	19.1
		NMS			*Other Nationals*	
At Work	84.3	85.3	85.6	55.7	54.9	55.7
(s.e)		(1.0)	(1.0)		(1.7)	(1.7)
Unemployed	8.8	5.7**	5.6**	12.7	6.5**	6.3**
(s.e)		(0.7)	(0.7)		(0.7)	(0.7)
Student	2.2	2.4	2.3	16.2	16.9	16.7
(s.e)		(0.5)	(0.5)		(1.3)	(1.4)
Home Duties	4.0	5.5**	5.4**	10.5	16.3**	16.1**
(s.e)		(0.6)	(0.6)		(1.0)	(1.0)
Other	0.7	1.1	1.0	4.9	5.4	5.1
(s.e)		(0.3)[s]	(0.3)[s]		(0.7)	(0.6)
SSE		13.1	14.0		73.5	72.6

Note:
(1) These estimates are for those aged 15 years and over.
(2) The Census figures are extracted from Table C0729 (*Census 2006* Published Tables).

Table A4: *Marital Status Distribution of Main Immigrant Groups in 2006 (%)*

Marital Status	*Census*	*Unweighted QNHS*	*Weighted QNHS*	*Census*	*Unweighted QNHS*	*Weighted QNHS*
		UK			*EU-13*	
Single	42.4	39.9	40.8	65.7	67.5	70.3*
(s.e)		(1.4)	(1.4)		(2.0)	(2.0)
Married	44.8	50.9**	50.5**	27.6	27.8	25.7
(s.e)		(1.4)	(1.4)		(1.9)	(1.9)
Divorced/ Separated	9.0	6.4**	6.1**	5.3	3.3*	2.9**
(s.e)		(0.6)	(0.6)		(0.7)	(0.6)
Widowed	3.8	2.8*	2.6**	1.4	1.4	1.1
(s.e)		(0.4)	(0.4)		(0.5)	(0.4)
SSE		51.2	44.9		7.3	30.6
		NMS			*Other Nationals*	
Single	62.0	66.0**	66.8**	49.2	52.0*	51.1
(s.e)		(1.4)	(1.4)		(1.3)	(1.4)
Married	31.9	30.4	29.9	45.8	44.6	45.6
(s.e)		(1.4)	(1.4)		(1.3)	(1.4)
Divorced/ Separated	5.5	3.2**	3.0**	3.8	2.4**	2.3**
(s.e)		(0.5)	(0.4)		(0.4)	(0.4)
Widowed	0.7	0.4	0.3	1.2	1.0	1.0
(s.e)		(0.2)	(0.1)		(0.2)	(0.2)
SSE		23.6	33.5		4.1	5.9

Note: The Census figures are extracted from Table C0440 (*Census 2006* Published Tables).

Table A5: *Highest Education Attainment Distribution of Immigrant Groups in 2006 (%)*

Education	*Census*	*Unweighted QNHS*	*Weighted QNHS*	*Census*	*Unweighted QNHS*	*Weighted QNHS*
		UK			*EU-13*	
Primary	8.0	16.0**	14.8**	4.0	3.0	2.4*
(s.e)		(1.1)	(1.1)		(0.8)[s]	(0.6)[s]
All Secondary	54.8	48.8**	48.6**	31.8	32.6	31.4
(s.e)		(1.5)	(1.5)		(2.3)	(2.4)
All Third Levels	37.2	35.2	36.7	64.2	64.4	66.2
(s.e)		(1.4)	(1.5)		(2.4)	(2.4)
SSE		104.0	84.9		1.7	6.7
		NMS			*Other Nationals*	
Primary	4.6	9.8**	9.0**	7.2	9.8**	9.3*
(s.e)		(1.2)	(1.2)		(1.0)	(1.0)
All Secondary	69.7	61.4**	60.7**	37.1	40.5	40.0
(s.e)		(1.9)	(1.9)		(1.7)	(1.8)
All Third Levels	25.7	28.8	30.3*	55.7	49.7*	50.7**
(s.e)		(1.7)	(1.8)		(1.8)	(1.9)
SSE		105.5	121.5		54.3	37.8

Note:
(1) These estimates are for those aged 15 years and over, and whose education has ceased.
(2) The Census figures are extracted from Table C1031 (*Census 2006* Published Tables).

APPENDIX B

2011 ESTIMATES

Table B1: *Sex Distribution of Main Immigrant Groups in 2011 (%)*

Sex	*Census*	*Unweighted QNHS*	*Weighted QNHS*	*Census*	*Unweighted QNHS*	*Weighted QNHS*
		UK			*EU-13*	
Male	50.2	50.7	50.3	48.3	43.4*	48.4
(s.e)		(1.4)	(1.4)		(2.2)	(2.3)
Female	49.8	49.3	49.7	51.7	56.6*	51.6
(s.e)		(1.4)	(1.4)		(2.2)	(2.3)
SSE		0.5	< 0.1		48.0	<0.1
		NMS			*Other Nationals*	
Male	50.3	50.4	50.3	49.8	48.1	49.9
(s.e)		(0.8)	(0.8)		(1.0)	(1.1)
Female	49.7	49.6	49.7	50.2	51.9	50.1
(s.e)		(0.8)	(0.8)		(1.0)	(1.1)
SSE		<0.1	0.0		5.8	<0.1

Note: The Census figures are extracted from Table CD620 (*Census 2011* Published Tables).

Table B2: *Age Distribution of Main Immigrant Groups in 2011 (%)*

Age Group	*Census*	*Unweighted QNHS*	*Weighted QNHS*	*Census*	*Unweighted QNHS*	*Weighted QNHS*
		UK			*EU-13*	
0-14	9.9	12.8*	9.9	8.9	11.3	8.8
(s.e)		(1.4)	(1.1)		(2.0)	(1.5)
15-24	10.6	9.9	10.3	11.8	13.4	13.2
(s.e)		(1.0)	(1.1)		(1.8)	(1.9)
25-34	12.2	9.2**	10.8	34.5	25.3**	27.4**
(s.e)		(0.9)	(1.1)		(2.4)	(2.5)
35-49	33.1	33.2	33.8	31.1	34.8	36.3*
(s.e)		(1.5)	(1.5)		(2.3)	(2.4)
50-64	21.5	20.8	20.9	9.8	10.2	9.4
(s.e)		(1.4)	(1.4)		(1.6)	(1.5)
65+	12.7	14.1	14.3	4.0	5.0	4.9
(s.e)		(1.2)	(1.3)		(1.1)[s]	(1.1)[s]
SSE		20.4	5.5		107.8	80.4
		NMS			*Other Nationals*	
0-14	16.8	15.3	16.8	15.9	16.7	15.9
(s.e)		(0.7)	(0.8)		(1.0)	(1.0)
15-24	12.2	12.9	12.5	14.6	13.7	13.5
(s.e)		(0.7)	(0.7)		(1.1)	(1.1)
25-34	45.3	43.0	44.2	32.0	28.4*	31.2
(s.e)		(1.2)	(1.2)		(1.4)	(1.5)
35-49	20.4	23.2*	21.4	30.5	33.4*	32.1
(s.e)		(1.0)	(0.9)		(1.2)	(1.3)
50-64	5.0	5.4	4.9	5.8	6.7	6.2
(s.e)		(0.6)	(0.5)		(0.7)	(0.7)
65+	0.3	0.2	0.2	1.1	1.1	1.0
(s.e)		(0.1)[s]	(0.1)[s]		(0.3)[s]	(0.3)[s]
SSE		16.0	2.3		23.6	4.6

Note: The Census figures are extracted from Table CD620 (*Census 2011* Published Tables).

Table B3: *Economic Status Distribution of Immigrant Groups 2011 (%)*

Economic Status	*Census*	*Unweighted QNHS*	*Weighted QNHS*	*Census*	*Unweighted QNHS*	*Weighted QNHS*
		UK			*EU-13*	
At Work	46.4	45.7	46.6	68.9	65.3	66.8
(s.e)		(1.7)	(1.8)		(2.6)	(2.5)
Unemployed	14.9	12.3*	11.9*	8.9	7.1	7.5
(s.e)		(1.1)	(1.1)		(1.5)^s	(1.5)^s
Student	8.2	8.1	8.1	9.6	11.2	11.0
(s.e)		(1.0)	(1.0)		(1.8)	(1.7)
Home Duties	10.6	13.6**	13.5**	6.5	11.0**	9.7*
(s.e)		(1.1)	(1.1)		(1.6)	(1.4)
Other	19.9	20.3	19.9	6.1	5.4	5.0
(s.e)		(1.4)	(1.4)		(1.1)^s	(1.1)^s
SSE		16.4	17.5		39.5	19.8
		NMS			*Other Nationals*	
At Work	67.1	64.4*	65.0	48.9	52.4*	52.4*
(s.e)		(1.2)	(1.2)		(1.6)	(1.6)
Unemployed	19.4	17.2*	17.1*	16.5	9.1**	9.3**
(s.e)		(0.9)	(0.9)		(0.8)	(0.9)
Student	5.5	6.5	6.0	20.4	18.0	18.1
(s.e)		(0.6)	(0.6)		(1.3)	(1.4)
Home Duties	6.0	10.1**	10.0**	9.3	16.9**	16.5**
(s.e)		(0.6)	(0.6)		(1.0)	(1.0)
Other	2.0	1.9	1.8	5.0	3.6*	3.6*
(s.e)		(0.3)^s	(0.3)^s		(0.5)	(0.5)
SSE		30.0	25.6		132.5	123.2

Note:
(1) These estimates are for those aged 15 years and over.
(2) The Census figures are extracted from Table CD306 (*Census 2011* Published Tables).

Table B4: *Marital Status Distribution of Main Immigrant Groups in 2011 (%)*

Marital Status	*Census*	*Unweighted QNHS*	*Weighted QNHS*	*Census*	*Unweighted QNHS*	*Weighted QNHS*
		UK			*EU-13*	
Single	39.4	39.1	37.8	60.9	55.2*	54.8*
(s.e)		(1.7)	(1.7)		(2.5)	(2.6)
Married	46.6	49.3	50.6*	32.0	38.0*	39.0**
(s.e)		(1.7)	(1.7)		(2.4)	(2.5)
Divorced/ Separated	9.9	8.3	8.3	5.7	5.0	4.4
(s.e)		(0.9)	(0.9)		(1.0)	(0.9)
Widowed	4.2	3.3	3.3	1.4	1.8	1.8
(s.e)		(0.6)	(0.6)		(0.6)	(0.6)
SSE		10.8	21.9		69.1	88.1
		NMS			*Other Nationals*	
Single	54.4	54.7	56.4	46.5	42.5**	42.8**
(s.e)		(1.1)	(1.1)		(1.3)	(1.3)
Married	37.7	39.9	38.4	48.1	53.3**	53.3**
(s.e)		(1.2)	(1.2)		(1.3)	(1.4)
Divorced/ Separated	7.0	4.6**	4.4**	4.4	3.4	3.1*
(s.e)		(0.5)	(0.5)		(0.5)	(0.5)
Widowed	0.8	0.7	0.7	1.0	0.8	0.8
(s.e)		(0.2)	(0.2)		(0.2)	(0.2)
SSE		10.7	11.3		44.1	42.5

Note: The Census figures are extracted from Table CD624 (*Census 2011* Published Tables).

Table B5: *Highest Education Attainment Distribution of Immigrant Groups in 2011 (%)*

Education	*Census*	*Unweighted QNHS*	*Weighted QNHS*	*Census*	*Unweighted QNHS*	*Weighted QNHS*
		UK			*EU-13*	
Primary	8.9	9.4	9.1	2.6	3.8	3.5
(s.e)		(1.0)	(1.0)		(1.0)[s]	(1.0)[s]
All Secondary	53.8	50.1**	49.7**	34.1	37.0	35.2
(s.e)		(1.8)	(1.9)		(2.7)	(2.8)
All Third Levels	37.3	40.6**	41.2**	63.3	59.2	61.3
(s.e)		(1.8)	(1.8)		(2.9)	(3.0)
SSE		24.8	32.1		26.7	6.0
		NMS			*Other Nationals*	
Primary	3.8	5.9	5.4	5.2	7.0	6.5
(s.e)		(0.7)	(0.7)		(0.9)	(0.8)
All Secondary	66.5	64.6	64.1	42.2	40.2	39.5
(s.e)		(1.4)	(1.4)		(1.7)	(1.7)
All Third Levels	29.7	29.5	30.5	56.3	52.8	54.0
(s.e)		(1.3)	(1.4)		(1.8)	(1.8)
SSE		8.1	9.0		19.5	14.3

Note:
(1) These estimates are for those aged 15 years and over, and whose education has ceased.
(2) The Census figures are extracted from Table CD922 (*Census 2011* Published Tables).