

# How much has the UK's overseas-born population actually contracted since the onset of Covid19 ? a research note

Ian Gordon, LSE,

[I.R.Gordon@lse.ac.uk](mailto:I.R.Gordon@lse.ac.uk)

21st March 2021

## 1. Introduction

Both nationally and internationally, the scale and awful impact of the Covid 19 pandemic has brought a range of new statistical initiatives focused on its dynamics, effects of public health responses to these and the challenges of mitigating its most significant socio-economic effects. But there have also been more diffuse side-effects on the basic infrastructure of statistical knowledge and monitoring capabilities on which modern societies rely.

In the UK case one example, hitting the headlines in early 2021, involved reports of a rather dramatic fall in the UK's population, of a million or so, during 2020, apparently as a result of many (more or less) recent migrants having 'gone home' in reaction to particular worries about Covid-induced risks. Heightened levels of overseas migration in recent decades had become a politically-charged matter in the UK, contributing significantly to a populist upsurge and to support for Brexit as a means of bringing this under control. Coverage of an apparently quite separate Covid19 reversal of the migration balance has so far focused on potential macro-economic effects (e.g. in relation to the March 2021 Budget) and possible labour supply or housing market impacts. These could be very sensible concerns, given the scale of the reported shift, but the data-base for all such reports is extremely fragile, because (in current circumstances) only a single survey source is available, which is neither designed to monitor aggregate population change nor immune to disruptive effects of social distancing on household-based data collection. Reporting of commentaries (notably by Sumption, 2021) which have questioned what can be inferred from this source is in some danger of playing into populist doubts about the credibility of 'official' reporting of migrant numbers, by suggesting that all is now completely uncertain. This research note seeks to fill the gap between these two journalistic responses, with a realistic estimate of the scale of change - and of the margins of ignorance, by distinguishing between the reliable and unreliable elements in the Labour Force Survey data - as well as of other possible ways of triangulating indications of very large changes.

This research note pursues the question of how large any decrease in the overseas-born population of the UK (and of major settlement regions) has really been, since the onset of the Covid19 pandemic last spring, in three stages. The *first* of these (in section 2 ) follows-up on O'Connor and Portes' (2021) highlighting of published ONS Labour Force Survey data that indicated a considerable reduction in this sub-population, between the summer quarters of 2019 and 2020, and their contention that the significance of this was largely missed by ONS because of a pre-fixed estimate of overall population change. The analysis in this section assumes, with them, that the reported reduction in overseas-born residents was real, and that the LFS sample data provide a reliable basis for estimating its size. The question at issue then is just how these sample numbers should be

grossed-up. My answer to that question involves a potential decrease in the non-UK-born population about mid-way between the ONS and the O'Connor/Portes estimates, and sharing their headline figure of 'around one million' - a historically unprecedented contraction, implying a truly massive reversal of the balance of UK migration.

The second stage (reported in section 3), focuses on the possibility that this 'loss' was actually a statistical illusion, attributable to a new source of non-response bias (within the sample data) arising from resort to a socially-distanced method of panel recruitment for the LFS during March 2020, in the face of surging Covid19 cases and lockdown restrictions. The credibility of this idea came from the ONS's own discovery of such a biasing effect in relation to housing tenure (ONS, 2020a), and Sumption's (2021) suggestion that some overseas-born migrants with less English-language fluency and confidence might be substantially less responsive to a phone-based rather than personal approach. My conclusion is that there is very strong prima facie evidence both for the emergence of such a bias and for that explanation of it – on a scale that seems to account for the major part (at least) of the reported fall.

The *third* stage (in section 4) then addresses the question whether the part remaining after initial correction for this effect - displaying an erratic time pattern, not readily linked to fluctuating pressures/opportunities for 'moves back home' - might reflect additional patterns of response bias at the actual survey stage (even though that was already phone-based). My conclusion here (based like the previous section on LFS data spanning the period January 2019 to November 2020) is that adjustment needed also to be made for the impact of the language-related factor on attrition of the LFS panels, with the effect of removing another large tranche of the apparent reduction in numbers of overseas-born residents. As a best estimate of real change in the size of the UK's overseas-born population this yielded a figure (235 thousand) closer to a quarter of a million than the initial headline estimates of a million, and with a different pattern of geographic concentration (more in the industrial regions of the English north/midlands rather than around London). But this has a rather large margin of error to it, considerable month-month variability in estimates, and real doubts as to how much of it is genuine since its composition is strongly skewed toward the same (linguistically challenged) sub-groups as were involved in the two identified forms of response bias.

A final reflective stage (in section 5) addresses the possibilities for more qualitative forms of triangulation to test the credibility of such large reported population shifts in a context where quantitative sources are more limited and/or less reliable.

## **2. Labour Force Survey Estimates of UK Population Change**

The Labour Force Survey is the most substantial household-based survey in the UK (and the basis for the Annual Population Survey). It is used extensively for both time-series and cross-sectional analyses involving the size of all kinds of population sub-groups, though it is not designed to provide estimates of the total population, overall or indeed by area and age. Grossing up of sample responses involves weighting factors incorporating **both** periodic cross-checks of *representativeness* against other data sources for particular sub-groupings (notably with the decennial Census), **representativeness and** overall population estimates for the nation and geographic areas, to get the

*scaling* right. The control totals used are the latest available, which ahead of the publication of official (administratively-based) mid-year estimates rely on trend projections from recent years.

This practice reflects (generally sensible) judgements about both relative margins of error for aggregate numbers from the two sources and about their degrees of fixity, i.e. the degree to which the size of sub-groups is constrained by inertia/changes operating at an aggregate level rather than capable of influencing these. Specifically, this might involve a judgement about whether a significant shock to the numbers of overseas-born in the UK would generate a balancing national level adjustment to the numbers of UK-born residents. At the scale of a city-region that seems quite possible, if not all at once (cf. Gordon and Champion, forthcoming). At the national level, however, it is pretty much inconceivable. But, as O'Connor/Portes (2021) pointed out, it is what the ONS-published LFS estimates for the period spanning the onset of the Covid pandemic implied – with a reported contraction of approaching a million overseas-born residents (after decades of expansion) being balanced statistically by a six-fold acceleration in the reported growth rate of the UK-born.

This was clearly an oversight, and not picked up as quickly as the one identified in relation to sudden major discontinuities in the mix of housing tenures (ONS, 2020a). As is recognised, that may not have been a specifically housing-related phenomenon – and the salient implications followed up were labour-market related. But ONS corrections were (reasonably) focused on the tenure aspect as one especially unlikely to show large short-term shifts<sup>1</sup>.

O'Connor/Portes (2021) propose a method of correcting for the failure to recognise that a large fall in the overseas-born population would directly affect the overall population change. This simultaneously enlarged the estimated fall in the non-UK-born population, replaced ONS's suggestion of a comparably large *increase* in the UK-born population with a small decrease, and also modified the regional distribution of overseas-born population losses (concentrating the majority of this on London). The thrust of these changes (or at least the first two) makes substantial sense – provided the underlying survey responses can be trusted. The specific adjustment formula proposed is rather opaque in its bases, however, specifically in terms of what is to replace ONS' baseline scaling assumption, namely projection of an established overall population growth trend.

My alternative version, combining:

- the spirit of the O'Connor/Portes critique – namely that this overall trend is not determined independently of fluctuations in overseas migration flows; with
- the ONS principle of simply scaling grossed-up survey estimates to some to some robustly estimable total,

just uses the (more stable) trend in *UK-born* numbers to scale those estimates. National (and regional) scaling/correction factors, are calculated from the ratio of externally estimated UK-born numbers to the total of individually-weighted (ONS PWT18) survey responses for that sub-population, and then applied to all respondents, wherever born.

Starting from a (4 year) trend-based estimate of modest growth in the UK-born population (of some 200 thousand over the year), this approach yields an estimated fall in the overseas-born population

---

<sup>1</sup> With an initially reported increase of about one quarter for wholly owned dwellings and a matching scale of decrease for rented properties.

midway between those of ONS and O'Connor/Portes (1.1 million over the year compared with 0.9 and 1.3 million, see Table 1). This, of course, presumes that the basic LFS survey counts for this sub-group are realistic. The implied overall UK population decline across the year is, however, much closer to O'Connor/Portes than ONS, whose starting assumption is one of continued growth, secured arithmetically by a substantially inflated estimate of the UK-born population, shown as increasing by 1.2 million against the 0.2 million or so of recent years<sup>2</sup>.

In any of these versions, the LFS data suggest that an absolutely enormous population shift occurred out of the UK during or between the lockdown phases last year. As a proportion of the overseas-born population, much of which is the outcome of great waves of in-movement during recent decades, the scale of this reported loss appears more modest – of the order of 10-15%. And it is possible to imagine that, when travel gateways reopen and normalcy returns, such a relatively marginal shift might get reversed in a few years. The real significance of this (reported) contraction, and the unlikelihood of easy reversal becomes much more evident, however, when a more disaggregated perspective is applied. This reveals firstly that rather than being diffused across the overseas population as a whole, the phenomenon has been heavily concentrated among particular sub-groups, where the impact would be far from marginal. And secondly it suggests that some apparently firmly settled types of household are involved, rather than just the more fluid, recently-arrived stereotype of young, single people who readily come-and-go and then come-back.

In particular, with my version of these estimates, it appears that net reductions would have been entirely concentrated amongst those members of the overseas-born population with relatively limited education (completed before the age of 21). More specifically, most (60%+) of the overall recorded reduction in the non-UK born population came from very large proportionate decreases (of a third or more) in numbers from a people with a combination of characteristics shared by only about one fifth of the non-UK born population. Specifically these were those:

- with modest educational qualifications (completed before age 21);
- arriving in the UK in 2000 or after (especially during 2009-2017); and
- coming from elsewhere in the EU (preponderantly, but not exclusively, the A10 East European accession countries); from South Asia; or a number of other groups of relatively poor countries (in eastern Europe, the middle east and North Africa, though not sub-Saharan Africa).

These pattern generalisations are consistent with (and indeed reflect) qualitative hypotheses (e.g. from O'Connor/Portes) about who might be expected to be both most vulnerable to negative impacts of the pandemic, and with less secure expectations about the longer term gains of staying put. In terms of regions of origins, there is broad consistency with a plausible tendency for probabilities of crisis-related return to be higher for migrants from relatively poor countries that are not too distant from the UK (within 2-3000 miles) – except for an apparently sizable South Asian element.

---

<sup>2</sup> The point should be registered that higher birth rates among an enlarged overseas-born population are a major positive factor in even this quite modest level of growth in the UK-born population. Emigration of families on the scale suggested by the LFS estimates of overseas-born population decline could quite substantially reduce that growth, partly because of births that might not occur in the UK as a result, but primarily because of the emigration of UK-born children accompanying their overseas-born parents.

Maybe, as we shall consider below, there are other ways in which the patterns of apparently concentrated reduction/return might be understood. And maybe also, the degree of concentration and implied speed of exodus are not just matters for real concern (e.g., in terms of labour market impacts) but also reasons for questioning whether it could credibly have been happening on this scale without coming to attention in other ways.

To address these questions we need to look more closely at how the LFS' data collection methods may have been affected by the pandemic situation in ways particularly affecting responses from some sub-groups within the overseas-born population.

### 3. Identifying Response Bias Effects of the Shift to a Distanced Survey Recruitment Regime

The UK Labour Force Survey is a very large and complex operation. But in essence it is a continuing household-based survey designed to yield reliable aggregate data for a very wide range of socio-economic variables on a quarterly basis – and data on individual level change over a 4 quarter interval. To achieve this it recruits panels of potential respondents for a sequence of 5 quarterly surveys (identified as *waves*), with a balance between those starting/finishing in each of the four quarters.

Sample responses are grossed on up on the basis of varying personal weights which reflect **both** evidence on differential achieved responses from those with different basic attributes (correcting for response bias) **and** on the required scaling to match known population aggregates (As discussed in the previous section). Evidence on patterns of response bias is generally retrospective, dependent on comparisons with other sources that are comprehensive and/or less reliant on tracing and voluntary collaboration by members of a diverse population. Exceptionally, however, it may come (as in the scaling case) from evidence of unexpected discontinuities in patterns that are more credibly explicable in relation to the working of the survey process than to 'real' socio-economic shocks.

In relation to Covid19 and the migrant population, one shock may clearly have had both kinds of effect. Perceived dangers of remaining in UK cities, possibly in overcrowded conditions without extended family support, or the realities of lost income from 'locked-down' activities, might well have encouraged some groups within the overseas-born population to 'return home' in search of security. But, on the survey front, a necessary shift in the mode of recruitment to the survey panel, to a socially-distanced telephone basis (a new *regime*, as I shall refer to it), may also very well have had a particularly negative effect on tracing and securing a positive response from particular groups within the overseas-born population.

Distinguishing these two sources of correlation between recognition of the covid19 pandemic in the UK and survey counts of the overseas-born population might in principle be made either on the basis of:

- the timing and dynamics of effects in relation to how the two processes might be expected to work in practice – when increased risks became evident and international mobility was more/less restricted *versus* when shifts in the recruitment regime would have impacted on actual survey responses; and

- the particular groups who might be expected to be most affected, whether actively in the case of 'going home' or more passively in that of not responding (at all, or positively) to attempted phone contacts - maybe those with least commitments/support in the UK *versus* those in unstable domestic situations and/or those least confident in dealing with unsolicited English-language phone approaches.

The timing issue is rather unclear, at least in relation to the processes conditioning actual return moves, including the two distinct waves of infection/lockdowns during 2020 and the periods in which international travel restrictions were most stringent. On the survey-process side there is a clearer break date, during March 2020, when the recruitment regime simply switched away from any face-to-face contact (ONS, 2020b). Any impacts of that specific change on levels/patterns of survey interview response would have been stepped in their significance – and indeed still be ongoing at the time of writing – since just one fifth of the panel actually gets changed each quarter. There might, of course, also be various other knock-on effects (both positive and negative) from 'real' external pandemic effects via domestic circumstances/staying home on a respondent's willingness to go through with an interview. But the recruitment regime switch is the one which should be quite clearly identifiable on a timing basis - given that the LFS database for each quarter distinguishes which wave a respondent is in – and the one linked to the clearest hypothesis about those most likely to be lost to the panel (in relation to confident English-language interactions).

To pursue this hypothesis, a combined QLFS data set spanning the period January 2019 to November 2020 was divided into three survey process groups:

1. a pre-Covid group for all responses with interview dates prior to a switching date in mid-March 2020;
2. an 'old regime' group from panels interviewed after that date, but who had been recruited before the change; and
3. a 'new regime' group identifiable as having been recruited (as well as interviewed) after that date.

For the 23 month period covered, these represented respectively 63%, 23% and 14% of the grossed-up responses (on my adjusted version of the ONS PWT18 weights).

As can be seen from Figure 1 (which identifies both of the first two groups with the 'old regime') the proportions of non-UK born respondents picked-up differ substantially between the two recruitment regimes, and much less clearly between the pre/post Covid outcomes from groups recruited before the March 2020 break. It certainly looks as if the shift in method would have been the key factor in the overall LFS finding of a substantial post-Covid drop in the size and population share of the overseas born part of the UK population.

To see how far this shift could be associated with introduction of a new response bias in terms of recognised language competences, a broader comparison was made between average mixes of respondents, not only in terms of the overall proportion born overseas but also between sub-groups differentiated in terms of their English language confidence. The basis of that distinction was a pair of LFS questions asking whether respondents had faced a language difficulty in education or in

getting/keeping a job<sup>3</sup>. These data could not be used directly at the individual level since they were last asked in summer 2018<sup>4</sup>. A broader case could be made anyhow - since the question is subjective and answers may be conditioned by factors ranging from mood to accidents of experience - for a less individualistic approach, focused instead on identifying socio-demographic groups (within the overseas-born population) whose members seem to face systematically higher or lower risks of experiencing such difficulties. Practically, the latter approach had to be adopted. This used a logistic regression analysis of the 2018 data to identify relatively durable variables which exerted strong influences on the probability of an individual experiencing (and reporting) such language-related difficulties. All overseas-born respondents in the 2019/20 data-set were then placed on a scale (and grouped) on the basis of the predicted probabilities for people like them.

The logit analysis initially identified 5 personal characteristics with significant differentiating effects on this risk: the overseas region of birthplace; ethnic origin; age of completion of education; years since arrival in the UK; and (for those from regions other than North America or Oceania) whether they had arrived in the UK before age 6. Subsequent analyses suggested an additional geographic factor in terms of region of residence within the UK, with a significantly lower incidence of reported difficulties in the south of England. Specific categories and related effect strengths are reported in Table 2, both without and with the regional variable, with the latter version being finally adopted.

On the latter basis, predicted probabilities of reporting such problems were computed for all overseas-born respondents in the 2019-20 data set, who were then grouped into quartiles<sup>5</sup>, with the majority of those vulnerable being concentrated in the first of these, as Table 3 indicates<sup>6</sup>.

A comparison (in Table 4) of the profiles of the three survey process groups, in terms of the proportions of the (grossed-up) respondents makes three things evident:

- across the full set of LFS respondents there was a drop of 1.4 percentage points (from 14.6% to 13.2%) in the recorded share of the overseas born population (compare rows 1 and 4 of Table 4) before/after March 2020. This is equivalent nationally to about 850 thousand people<sup>7</sup>, or 9.6% of the base numbers in this nativity group; but
- in relation to the comparable response set from the old-regime element of the post-March surveys, a more realistic estimate of the drop in share seems to be 0.6 percentage points, to 14.0% (compare rows 1 and 2). This is equivalent to about 380 thousand people, or 4.3% of their base numbers; and
- consistent with the hypothesis of reduced recruitment rates from those with language communication difficulties, the proportion of overseas-born residents recorded by the new

---

<sup>3</sup> The question was actually asked of all those in households with a non-UK first language, but has here been applied just to those born outside the UK (ignoring non-migrant members of households with non-British first languages, and assuming those outside such households have no comparable language difficulties).

<sup>4</sup> And should next be in summer 2021

<sup>5</sup> The quartile breakdown used here, and subsequently, was chosen after experiments with a finer disaggregation, which suggested a break at around the 25% mark between the groups most susceptible to difficulties with English language use and the majority of the overseas-born population.

<sup>6</sup> This observation follows from a statistical expectation that those in the top quartile face more than double the average chance (among the overseas-born) of experiencing/reporting such problems (19% against 8%).

<sup>7</sup> These numbers, and subsequent numerical estimates, with the underpinning analysis, all relate to those under the age of 75, on the premise that above this age changes are very unlikely to reflect international migration, which is the focus of interest.

regime element is much lower, at 11.9%, with a 2.1 percentage point shortfall from that in the old regime estimate (compare rows 2 and 3). This is concentrated in the two quartiles involving those with a higher incidence of these problems. In the top quartile it seems that one third of the relevant sub-population (1.1% out of 3.4%) was effectively lost to the panel when the recruitment method had to be switched, while in the second quartile it was about one fifth (0.8% out of 3.7%).

For the purposes of this note at least, the obvious way of getting around this particular source of downward bias in post-covid estimates of the overseas-born population is by just putting the new regime element entirely on one side, and looking solely at those recruited to the panel before mid-March 2020.

Fortunately, because of how the LFS is designed this still leaves a large number of respondents sampled on the same basis as those for the pre-Covid years. Focusing on these more than halves the implied scale of population shrinkage (and presumed return migration) among the overseas-born, though in absolute terms the estimate remains unprecedentedly large, at 380 thousand (as noted above). Perhaps surprisingly, this cut-down estimate still shows a very uneven distribution across the language-confidence distribution. Indeed an even more concentrated one, with two thirds of the remaining gap – and the only statistically significant element – seemingly concentrated among those in the top quartile in terms of potential language problems, with an implied fall in numbers of about one tenth (compare lines 1 and 2 of the Table). Whether that might reflect some other new source of response bias or a ‘real’ influence of other associated factors remains to be unpicked.

#### **4. Response Bias Effects in Later Waves of the Survey**

Differential patterns and degrees of bias can emerge, *within* the recruitment panel (not just at the point of recruitment), as response rates are attenuated - by residential movement, changes of circumstance and/or boredom – in ways that may well differ between population groups, maybe reinforcing those operating at the panel recruitment stage. And they might be generally exaggerated by the circumstances of the pandemic and lockdown.

Only the last of these would seem relevant to the question of how much of the recorded fall in the migrant population is real/spurious, because the staggering of panel recruitment means that in any quarter there should be a consistent mix of respondents in their 1<sup>st</sup> through to 5<sup>th</sup> rounds of surveys. However, when (as in the last section’s analyses) a distinction is made between those recruited under new/old ‘regimes’ that ceases to be the case. The concern of this section is thus with two different ways in which response bias effects cause might depress the proportion of overseas-born respondents within the ‘old-regime’ panellists at some points after March 2020: through

- the circumstances of the crisis exaggerating a particular tendency for overseas-born respondents with more language difficulties to drop out after earlier waves of participation; and/or
- a combination of established wave-specific biases plus a changing wave-mix among the remaining ‘old regime’ panellists after March 2020

And with trying to distinguish these from time dependent-influences on responses because of exit from the UK population – though the latter would not be separable from effects on willingness/capacity to participate.

Statistically the approach is to exclude the ‘new regime’ respondents and then look for statistically significant patterns of variance in the proportions of UK-born respondents and those in ‘language differences in the probability of responses by people from different groups (the UK-born and the four language difficulty quartiles of the overseas born). In practice this was carried out through a series of dummy variable regressions with the membership of the UK-born population, the overseas-born, and the language difficulty quartiles as dependent variables and independent variables relating to survey date and interview wave. In addition to the basic ONS personal weights, responses were scaled to total responses in each month, so that coefficients represented effects on percentage shares in the total (recorded) population<sup>8</sup>.

A series of experimental analyses of the monthly series, from January 2019 on, all confirmed the significance of inter-wave variations in the recorded shares of the overseas-born (and particularly of the sub-groups more likely to lack confidence in English language use). But they also all struggled to find significant temporal effects that might be identified associated with either form of Covid impacts. Two reasons, apart from the varying ways in which these might have been realised over time, are the relatively short period in which movement effects might have been anticipated and the degree of apparently random month-to-month variability in the series (see Figure 1). Levels of general public concern about the pandemic threat appeared to grow very sharply during March (Ipsos/MORI, 2020). But opportunities for international travel were very restricted during the 2 months that followed. Heathrow passenger data suggest indicate traffic as starting to pick up again during June and for EU destinations reaching almost a third of 2019 levels during Explorations homed in on a single comparison between the period from June-November 2020 and the early period, though this implied a much more concentrated response around mid-year than might reasonably have been expected.

Results from this analysis are presented in two Tables, focused on different aspects. The first (Table 5) of these simply shows how the scale and composition of response varied between waves - for the period up to March 2020 before the new regime came in and altered the inter-wave balance for the ‘old regime’ cases on which we focus. Essentially this shows that the (expected) attrition of response between waves was especially concentrated among the most linguistically challenged of the overseas-born groups. On top of an average pattern implying a drop-out rate of 30% between waves 1 and 5, it appears that there was an additional drop of 23% for those in the quartile of the overseas-born most vulnerable to language-related difficulties. There is some reason to believe that drop-out rates between waves will have increased during the pandemic (ONS, 2020a/b). But no evidence was found that this significantly exacerbated the tendency for these groups to get selected-out of later waves.

The second focuses on the question of how much and how significantly the size (and specifically the share) of the overseas-born population (and sub-groups within it) might have reduced in the six

---

<sup>8</sup> Simple linear regressions were preferred to logistic ones in this case because they enabled simple additions of coefficients across sub-groups.

months after May 2020<sup>9</sup>. Table 6 presents results from simple time series regressions of averaged shares within the (comparable) ‘old regime’ component of LFS respondents, in terms of the coefficient on a break dummy – in two versions, with and without controls for the wave effect. These suggest that potential effects are concentrated within the most disadvantaged sub-groups, and even there are not always of clear statistical significance, even when the numbers involved are far from trivial.

In overall numerical terms, a baseline result (ignoring inter-wave effects), reported in row 2 of Table 6) suggests a possible reduction of some 476 thousand in the UK’s overseas-born population (an estimated shift of 0.79 percentage points in its share of a 60.5 mn. population base) before/after end-May 2020. [This is about a quarter higher than that suggested by the analysis in the previous section, though the break dates differ by 2.5 months<sup>10</sup>]. However, as row 3 shows, after correction for the shifting Wave-mix (after March) within the ‘old regime’ respondent set on which the analysis focuses, this covid-impact estimate is *halved*, and comes down to some 235 thousand (i.e. 0.39 % of the same population base). In each case the standard error of estimate is about 75 thousand, so both estimates are still statistically significant at the 1% level. But the magnitude of the corrected estimate is *very* much smaller than the 3 face-value LFS estimates cited in section 2 – *between a quarter and a fifth of what had been suggested* (74% lower than the published ONS figure and 82% lower than O’Connor/Portes’ revised version). And, though there are differences in the time periods to which they refer, this is essentially a reflection of the degree of response bias introduced by the forced switch to a new recruitment panel method for the LFS in mid-March 2020.

Significantly, the effects seem to be concentrated entirely within either the top two (in the unadjusted version) or the topmost (in the corrected one) of the quartiles of predicted language difficulty among the overseas-born population. In the unadjusted case this is what we might expect given that the differential inter-wave effect was known to be concentrated there. But to find the adjusted effect to be even more concentrated is striking, since though the estimated effect there is not statistically significant (at the 5% level) it is important in relation to the size of this sub-group, with an implied 10% shrinkage after May 2020 (relative to the previous average).

It is not unreasonable to be suspicious that this concentrated effect might reflect some other ways in which representation of this group has shrunk during the crisis – implying a still more modest reduction in the size of the overseas-born population. But there is really no basis on which LFS evidence can be used to test this.

Other aspects of the composition and distribution of the implied 235 thousand fall are worthy of note. Some of these involve attributes such as country of origin and ethnicity which are clearly associated with the imputed language difficulty, while others relate to regions of residence within the UK, which had a less obvious association (Table 6). In terms of countries of origin it is notable

---

<sup>9</sup> Just to be clear, the comparison involves differences between *levels* in the months before and after the suggested break-point. The estimates thus relate to a posited step-change, rather than a dynamic adjustment via a continuing net outward flow. The latter seems more credible a priori but there is nothing in the data series up to November to suggest (or allow statistical description of) such a pattern.

<sup>10</sup> This reflects a difference in focus. The mid-March date used in the earlier analysis related to the lock-down point of change in survey procedures, while the end-May reference date for this analysis was based on a judgement as to the most plausible point for real decreases to have started (in relation both to external factors and graphing of the population series).

that there seem to have been no net reductions among those coming from the 'rich country' groups (EU15, North America or Australasia) – nor indeed from the more mixed 'rest of the world'. The birth regions showing sizeable reductions are in the rest of the EU, the contiguous regions (North Africa, the Middle East/Central Asia and Other Europe) and, more surprisingly South Asia). In relation to regions of residence within the UK a significant net loss is shown for London, but not on anything like the scale implied by the originally published data (Table 1), and very much less than is suggested for the English Midlands and North (particularly the North East, Yorkshire/Humberside and the West Midlands). Why that pattern should have emerged is another, substantively interesting research question to be pursued – not least because it suggests a new dimension to the UK's uneven regional geography

## 5. Concluding Observation

One of the significant side-effects of the Covid19 pandemic within the UK has involved challenges to the maintenance and consistency of established data series on basic demographic and economic matters. The field of international migration is an important case, since the suspension/demise of the International Passenger Survey has increased dependence on a Labour Force Survey which is not really designed to produce population estimates.

This case has been brought to the fore because published results for the size of the overseas-born population in summer 2020 pointed to a truly extraordinary scale of reversal in the balance of net migration - though the potential implications of that for the overall population (and hence labour supply, taxable capacity etc.) were obscured by prior assumption about the trend in this.

Its greater significance may be in exposing the vulnerability of even as established a survey as this to circumstances requiring a switch in practices (specifically here to a purely phone-based recruitment). Following up the suggestions by Sumption (2021) this note has used evidence from within the LFS's own data-base to show that this had a massive effect on estimates of the size of the post-covid overseas-born population, leading to initial estimates of reduction in its numbers which were inflated by a factor of about 4. A revised estimate of a reduction of some 235 thousand (still a remarkable number) may still conceal some inflation by other forms of selection bias that cannot be appraised purely from LFS sources.

There are lessons to be learned from the original estimates and reactions to these, since what made them initially credible, as well as remarkable, was the degree which they involved concentrations among specific sub-groups who might have been expected to be more likely to respond by returning 'home' - as well as others for whom this always seemed unlikely (such as the South Asians). But, what this should also mean is that there is a potential for cross-checking (triangulating) the credibility of some macro-change estimates against qualitative indicators for specific communities where big shifts are implied – to see whether community leaders or local businesses are aware of substantial changes or whether these are consistent with silences in the local press (e.g. about any very big reductions in Inner London's Bangladeshi community such as the unadjusted LFS data was suggesting). Publication of the reports earlier this year about very large falls in the UK's overseas-born population was accompanied by familiar *vox pop* illustrations/explanations in the press, fleshing out some trends which may not really have occurred. But it should be possible for

researchers to use the same kind of material (and its relative absence) to critically evaluate such reports, in a context where familiar statistical sources may have been thrown awry.

## References

Gordon, I.R. and Champion, A.G. (forthcoming, 2021) *The Dynamics of Migration Flows and Population Shifts in England's Metropolitan Super-region*.

Ipsos MORI (2020) *Coronavirus: Tracking UK Public Perception*, infographic snapshot, 6<sup>th</sup> November 2020 ([www.ipsos.com/ipsos-mori/en-uk/public-opinion-covid-19-coronavirus-pandemic#infographic](http://www.ipsos.com/ipsos-mori/en-uk/public-opinion-covid-19-coronavirus-pandemic#infographic))

Office of National Statistics (2020a) 'Coronavirus and its impact on the Labour Force Survey', 13<sup>th</sup> October, 2020 (Bob Watson)  
(<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/coronavirusanditsimpactonthelabourforcesurvey/2020-10-13#overview-of-the-introduction-o%E2%80%A6>)

Office of National Statistics (2020b) 'Labour Force Survey performance and quality monitoring report: July to September 2020', November ?, 2020  
(<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyperformanceandqualitymonitoringreports/labourforcesurveyperform...>)

O'Connor, M. and Portes, J. (2021) *Estimating the UK Population During the Pandemic*, Economic Statistics Centre of Excellence, 14<sup>th</sup> January 2021.  
(<https://www.escie.ac.uk/estimating-the-uk-population-during-the-pandemic/>)

Sumption, M. (2021). *Where did all the migrants go? Migration data during the pandemic*. Migration Observatory commentary, COMPAS, University of Oxford  
( [www.migrationobservatory.ox.ac.uk](http://www.migrationobservatory.ox.ac.uk))

## Acknowledgement

All data used here are from Quarterly Labour Force Survey data files downloaded (in February 2021) from the UK Data Archive at the University of Essex.

## Tables and Figure

**Table 1: Alternative LFS-based Estimates of Population Change between summer 2019 and 2020 by Nativity and Region – all assuming unbiased comparison of pre/post Covid situation**

Estimated Population Change Summer2019-Summer 2020 (000s)									
	UK Born			Non UK-Born			Total Change		
	ONS_PWT18	Portes-O'Connor	Gordon	ONS_PWT18	Portes-O'Connor	Gordon	ONS_PWT18	Portes-O'Connor	Gordon
North East	35	-7	5	-30	-14	-31	5	-20	-26
North West	54	-31	25	-33	-47	-35	21	-78	-10
Yorkshire and Humberside		-3	14						-110
East Midlands	133	2	10	-115	-128	-124	18	-130	
West Midlands	142	57	8	-116	-129	-129	27	-127	-119
East of England	256	..	28	-226	-254	-254	31	-197	-246
London	82	..	..	-37	..	-44	45	..	-16
South East	435	-32	53	-355	-669	-534	80	-701	-481
South West	99	..	33	-39	..	-49	60	..	-16
Wales	39	11	21	1	-7	-1	40	4	20
Scotland	3	25	3	7	11	6	9	36	9
Northern Ireland	-22	-48	0	41	41	43	18	-7	43
UK	0	-1	7	8	10	9	8	9	16
[East + South East ]	1256	-56	207	-894	-1294	-1143	362	-1349	-936
	181	-29	61	-76	-108	-93	105	-138	-32

**Sources:** ONS *Quarterly Labour Force Survey*, July-September quarter 2018.

**Notes:** The 'ONS' estimates are based on the PWT18 weights from the data files, without any adjustment; the other two apply adjustment/reweighting methods as described in the text, and for Portes/O'Connor published in their cited January 2021 Discussion Paper.

**Table 2: Logistic Regression of Reported Language Difficulty on Personal Attributes of Overseas-born Residents**

Variable	Personal Attributes Only (1)		Plus Region of Residence (2)	
	coefficient	t statistic	coefficient	t statistic
<b>Constant</b>	-3.919	12.5***	3.747	12.1***
<b>Years Resident in UK</b>	-.041	8.9***	-0.038	10.0***
<b>Overseas Region of Origin</b>				
EU (excluding A8 and A2)	0.342	1.4	0.320	1.4
EU8, EU2 and Other Europe	1.495	7.1 ***	1.443	7.1***
North America and Oceania (aka 'Anglo regions')	-21.598	99.9 ***	21.597	99.9***
Middle East, Central Asia, East Asia and North Africa	1.474	7.7 ***	1.391	7.7***
South Asia, and Central/South America	0.685	3.9 ***	0.690	4.1***
South East Asia and Sub Saharan Africa	0.0	..	0.0	..
<b>Came to UK before age 6</b> (from non-Anglo region)	-1.997	3.9 ***	-1.847	4.0***
<b>Age that Completed Education</b>				
none	1.975	5.2 ***	1.926	5.1***
16 or under	1.692	6.8***	1.580	6.4***
17-20	0.920	3.8***	0.934	3.8***
21 and over	0.458	1.9	0.494	2.0*
Still in education (16+)	0.0	..	0.0	..
<b>Ethnic Origin</b>				
White	0.0	..	0.0	0.0
Black	0.274	1.2	0.358	1.6
Indian, Pakistani, Chinese, Mixed/Multiple and Other	0.630	4.1***	0.637	4.2***
Bangladeshi and 'Other Asian'	1.153	6.1***	1.196	6.5***
<b>Region of Residence in UK</b> South of England			-0.380	4.6***
N	9659			
Pseudo R-squared (Nagelkerke)	0.175		0.180	

**Source:** ONS Quarterly Labour Force Survey, July-September quarter 2018.

**Notes:** 1. The dependent variable relates to overseas-born members of the UK population and takes the value 1 for those reporting a language difficulty, in education or in finding/keeping a job, and zero otherwise; 2. The South of England comprises the London, the Eastern, South East and South West regions; 3. Asterisks denote significance levels: \* = 5%; \*\*=1%; \*\*\*=0.1%.

**Table 3: Quartile Grouping of Overseas-Born Respondents by Predicted probabilities of Language Difficulty**

	High LDQ1	Moderate LDQ2	Minor LDQ3	Very Low LDQ4	ALL Overseas Born
<b>Probability Mean</b>	18.9%	8.1%	3.6%	0.7%	7.8%
Minimum	11%	5%	2%	0%	0%
Maximum	61%	11%	5%	2%	61%

**Source:** Quarterly Labour Force Survey data for January 2019-mid-March 2020.

**Notes:** 1. Probability estimates are based on the coefficients reported in col 2 of Table 2 from the summer quarter 2018 analysis; 2. In this table, and all that follow, analysis has been restricted to those under the age of 75.

**Table 4 Representation of Language Difficulty Quantiles by LFS Survey Groups**

Survey Group	Overseas Born					UK Born
	Probabilities of Language Difficulty				All	
	High LDQ1	Moderate LDQ2	Minor LDQ3	Very Low LDQ4		
Jan 2019-Mar 2020: All	3.79%	3.77%	3.63%	3.42%	14.61%	85.39%
From March 2020						
Old regime	3.38%	3.71%	3.54%	3.34%	13.98%	86.02%
New Regime	2.26%	2.95%	3.31%	3.40%	11.91%	88.09%
All (Old + New)	2.96%	3.42%	3.45%	3.37%	13.21%	86.79%

**Source:** Quarterly Labour Force Survey data for January 2019-November 2020.

**Note:** percentages relate to numbers grossed up on the basis of ONS2018 weights re-scaled for consistency (on a month-by-month basis) with trend numbers of the UK-born. No allowance is made for any possible shifts in the LFS coverage of the overseas-born

**Table 5. Selectivity in Attrition of Response by Wave**

	Share Relative to Total Responses						UK Born	Absolute Total Respondents
	Overseas Born							
	High LDQ1	Moderate LDQ2	Minor LDQ3	V. Low LDQ4	Not high LDQ2to4	All		
<b>Wave 1</b>	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
<b>Wave 2</b>	92.5%	100.1%	103.3%	103.3%	102.1%	99.4%	100.1%	101.6%
<b>Wave 3</b>	84.6%	96.8%	100.3%	105.9%	100.7%	96.1%	100.8%	84.4%
<b>Wave 4</b>	83.8%	93.9%	97.6%	107.9%	99.4%	94.9%	100.9%	75.5%
<b>Wave 5</b>	77.1%	91.7%	95.4%	106.6%	97.5%	91.7%	101.6%	70.0%

**Source:** Quarterly Labour Force Survey data for January 2019-November 2020.

**Notes:** 1. The estimates derive from regressions (reported in the next table) using individual data from the QLFS, dependent variables taking the value 1 for members of the groups concerned and zero for others, scaled to represent percentages of all respondents in the month concerned; 2. Indexes relative to initial wave responses are calculated from dummy variables for the wave concerned, with a control for observations after May 2020; 3. Observation from panels recruited under the (phone-based) ‘new regime’ have been excluded. The final column for total response numbers relates to months up to March 2020 only; 4. regression weights reflect the original ONS personal weights, appropriately rescaled for this analysis.

**Table 6: Estimated Shifts post-June 2020 in the Population Shares of the Overseas Born (overall and for groups with more language difficulty) from regressions with 2019/20 monthly data**

	Overseas Born				
	By Degree of Language Difficulty				All
	High LDQ1	Moderate LDQ2	Minor LDQ3	Very Low LDQ4	
Mean Population Share	3.62%	3.74%	3.62%	3.40%	14.37%
Estimated change post May 2020 Without Wave controls	-0.621%***	-0.124%*	0.011%	-0.054%	-0.788%***
With wave controls	-0.343%***	-0.007%	0.090%	-0.129%*	-0.389%**
N	509,958				

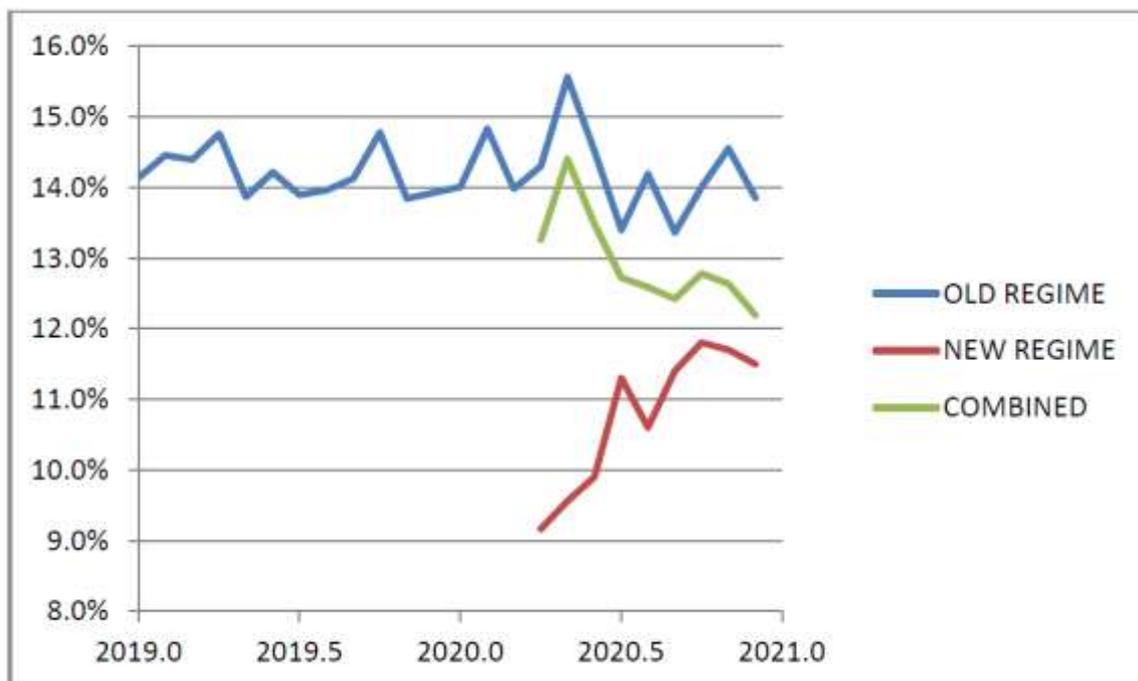
**Source:** Quarterly Labour Force Survey data files from UKDA

**Notes:** as for Table 5, excluding the reference to its total response column.

**Table 6: Final Estimates of Covid-Related Change in the UK’s Overseas- born Population : distribution by regions of birth and of UK residence.**

	Net Change in Numbers Born in these Regions (000s)		Net Change in Overseas Born Residents of these Regions (000s)
EU 15	19	London	-41
Rest of EU	-61	Rest of Southern England	+63
Adjoining regions (Other Europe, North Africa, Middle East and Central Asia)	-122	North and Midlands	-248
South Asia	-165	Wales, Scotland, Northern Ireland	+9
North America and Oceania	23		
Rest of the World	46		
Total (cf estimated UK total)	-260 (-235)		-217(-235)

**Figure 1: Estimated Share of the Overseas Born in UK Population by Survey Components**



**Source:** calculations from Quarterly Labour Force Survey data-sets.

**Note:** The low starting point and strong upward climb for the New Regime series during the first two quarters after the switch seems to reflect particular implementation problems at its outset in mobilising groups with more substantial language problems.