What to believe about the econometric models in three studies of the fiscal effects of immigration to the UK

Mervyn Stone
April 2015

On 14th April with 24 days to the Election, the small refurbished attic of the (Shepherd's) Bush Theatre was the venue for the Fabian Society launch of the pamphlet 'PLACES TO BE: Green Spaces for Active Citizenship'. Purporting to be an active citizen, I got to the meeting with the aid of a neighbour to hear DEFRA shadow minister, Maria Eagle, give a nicely-balanced account of how she would handle her environmental portfolio when (which is how she put it) she came out of the shadows.

Perhaps concerned lest my mental state matched a manifest physical infirmity, chairwoman Sally Gimson was reluctant to let me put the question to a panel that included the Labour candidate for Hammersmith, Andrew Slaughter. My question was whether (and if not, why not) the distributed pamphlet made any mention of population as a factor putting housing pressure on green spaces. Questions were dealt with three at a time, so I can't be sure that I heard Mr Slaughter saying that the population of some London borough had not increased since the 1930s—I had introduced myself as the widower of someone who had been Fabian secretary for a London borough. In fact, 'PLACES TO BE' does have five mentions of 'population', but none of them are to the politically sensitive expansion of population in the Blair & Brown years—an embarrassment that Candidate Slaughter may have been studiously evading.

The nearer we get to the Election, the less tolerance there seems to be in any public debate for any reference to that powerful feeder of population growth—the aggregation of the annual net immigrations so far constituted. Some of the debate took place in Civitas, and it did so rather controversially. That was when I responded critically in Stone (2013) to the study of the fiscal benefit of immigration by Professors Dustmann and Frattini in D&F (2013) which succeeded the earlier study Dustmann et al (2010). The controversy may have ended with CReAM's response in D&F (2014a) to the press release that accompanied my critique of D&F (2013) and with my declining to be provoked by the response's pretence that only the press release merited serious rebuttal—and that, in effect, the critique could be lightly brushed aside.

The present note has, however, been justifiably provoked by the publication of a third study D&F (2014b) in The Economic Journal. Although this publication is little more than a reformulation of D&F (2013), it deceived its readers in two respects—by not explaining how its significant changes related to the earlier studies and by remaining silent about the important elements of the controversy between D&F (2013) and Stone (2013). It was as if the Economic Journal too were complicit in an obfuscatory silence.

There is a widespread media-level view that Professors Dustmann & Frattini have said the last word on the fiscal benefits issue. This note disputes that view by examining in greater detail than did Stone (2013), the technical objections to a major component of the D&F case—namely, the econometric-based estimates of a 'differential' in the probability of receipt of state benefits between immigrants and natives. The D&F publications come from CReAM (Centre for Research and Analysis of Migration)—a richly-staffed group of researchers in the Department of Economics of University College London. The work of CReAM has been richly endowed since 2007 as a 'project' of the Anglo-German Foundation charity—a double richness of support has gone sadly unrewarded. The rest of this note will show that the quality of CReAM's research on at least one component of the project has belied the reputation of the supporting institutions. Without the controversy generated by this fall from grace, public discourse about the fiscal benefit of immigration could have focussed without distraction on the broader analysis of the net fiscal balance component of the issue in Professor Rowthorn's book (Rowthorn, 2014b).

The rich and complex provenance of the final publication

The publication of Dustmann and Frattini (2014b) by the Economic Journal is the latest of a sequence of three reports on the same theme from the Centre for Research & Analysis of Migration (CReAM) at University College London. Table 1 exhibits the subject matter of this note:

Table 1. Data, statistical method and 'differential in probability' estimates in three publications

Publication:	**Pub 1 (D&F & Halls, 2010)**		**Pub 2 (D&F, 2013)**		**Pub 3 (D&F, 2014b)**	
Immigrants questioned by the Labour Force Survey (LFS[a]) data-base:	A8 immigrants 2005/06 to 2008/09 (16 year-quarter samples)		LFS: All immigrants 2000/01 to 2010/11 (44 year-quarter samples)		LFS: All immigrants 2000/01 to 2010/11 (44 year-quarter samples)	
Statistical method for estimating the probability of receipt of benefits:	By 'linear probability' and least-squares		By 'linear probability' and least-squares		By 'normality-probit' and maximum likelihood?	
Estimate of the immigrant probability minus estimate of native probability (and its standard error):	Model 1 −23.3% [b] (0.7%)	Model 2 − 5.1% [c] (0.6%)	Model 1 −17.8% [b] (0.1%)	Model 2 − 8.4% [d] (0.1%)	Model 1 −17.4% [b] (0.2%)	Model 2 −15.5% [e] (0.2%)
Number of observations:	1,455,111	1,260,339	3,495,478	3,495,478	3,451,264	3,451,264

[a] LFS is a survey of working-age individuals in quarterly random samples of households.
[b] With year-quarters (as covariates to the immigrant/native status variable).
[c] With year-quarters, sex, education levels, a quadratic in age, a dummy for whether-or-not there are dependent children in an individual's household, and the number of those dependent children as covariates (23 in number?).
[d] With year-quarters, sex, and a quadratic in age as covariates (47 in number).
[e] With year-quarters, 5-year age-bands as covariates for each sex (75 in number?).

Pinning down the key figures of Table 1—e.g. the Model 1 percentage differences and the progression of the native/immigrant differences 5.1%, 8.4% and 15.5%, and why there are millions of observations—will be an unavoidably technical matter. It has to be done—if only because D&F rightly take the question *"Is a randomly drawn immigrant more likely to receive benefits…than a randomly drawn native …?"* to be *important for assessing immigration's fiscal cost* (Pub 2 p.10). Taking one step at a time, we start with a description of the source of the data that D&F rely on—the UK Labour Force Survey.

Structure of the Labour Force Survey

The Office of National Statistics paper (Browne and Alstrup, 2006) is so clear that it is better quoted than reworded:

> *The survey is based on a random sample throughout the whole of the United Kingdom. Every three months almost 53 thousand households take part in the survey. ... The survey collects information about the personal circumstances and work of everyone living in these households. ... We attempt to interview people who take part in the survey five times in all. An interviewer almost always visits their home to conduct their first interview. ... the interviewer then seeks the household's cooperation to take part in subsequent follow up quarterly telephone interviews, since measuring change over time in the household's labour market participation is very important. The other interviews take place at three-month intervals. In most cases we conduct these follow-up (the second, or third, and so on) interviews over the telephone. LFS interviewers try to collect information from each member of the household. Sometimes this is not possible—for example, if only one member of the household is at home when the interviewer calls. In such cases LFS interviewers collect information about absent members of the household from the (adult) person who they make contact with. ... Occasionally, interviewers cannot get in touch with anyone in the household, to conduct a follow-up interview. In this case we use the information that they provided at their previous interview. ... Much of the data collected on the LFS is on the basis of self-classification ... self-classification can lead to differences between the LFS and other data sources ... questions that require a familiarity with administrative systems (such as the benefits system…).*

Table 2. Face-to-face interviews 'S- ' and follow-up telephone interviews 'f ' of 44 quarterly household samples. Quarters were seasonal before 2006 but financial after a supposedly smooth transition.

2000/01	**2001/02**	**2002 to 2010**	**2010/11**	**2011/12**
Q1 Q2 Q3 Q4	Q5 Q6 Q7 Q8	Q9 ... Q40	Q41 Q42 Q43 Q44	Q45 Q46 Q47 Q48
S1 f f f	f S6 f f	f ... f	S41 f f f	F
* S2 f f	f f S7 f	f ... f	f S42 f f	f f
* * S3 f	f f f S8	f ... f	f f S43 f	f f f
* * * S4	f f f f	S9 ... f	f f f S44	f f f f
* * * *	S5 f f f	f ... S40	f f f f	

Table 2 displays what I take to be the framework of the data for Pub 2 & Pub 3. For each 'wave' of five interviews (S f f f f), there are LFS records for individuals interviewed in the initial home visits to about 10,000 or so households, and records for the absent household members those interviewed can speak for. There are also

records elicited for the first time in follow-up interviews, as well as records carried forward from quarter to quarter either without change or else updated if there is any change.

The D&F 'observations' are replicates of LFS records on the same individuals in successive quarters

With a household participation rate of about 80%, the 44 samples of Table 2 would have yielded about 420,000 households for interview. For 3.5 million 'observations', that makes about 8 observations per household. Neither Pub 2 nor Pub 4 give good reason for replicating records for the same individual (ONS has to do that as a statutory duty). Pub 2 simply says that to

> *increase the sample size, we pool the four quarterly waves in every fiscal year, which in the UK begins in April. Hence, for fiscal year t, we pool LFS quarters 2, 3 and 4 of year t and quarter 1 of year t+1 ...*

(which happens to ignore the change from seasonal to financial-year). Pub4 gives the reason as:

> *Information on receipt of state benefits, however, has been available only since 1998, whereas information on social housing exists for all the years considered in our analysis. We,* ***therefore,*** *increase the sample size by pooling the four quarterly waves in every fiscal year, which in the UK begins in April …* (my emphasis)

Whatever the reason for pooling, it remains that increasing the 'sample size' to millions has left room for bias between individuals of different status within households—between those who were always available for interview and those whose first interview came later in the wave and therefore had fewer carried-forward records. Such bias could well differ between immigrants and natives and it seems that CReAM may not have considered it. Pub 3 had the option of moving to a more appropriate use of the individual records underlying Table 2, in line with its straightforward intention to

> *assess the degree of welfare dependency of immigrants relative to natives based on the responses in each LFS wave on whether individuals receive state benefits … .* (p.22)

Since responses may have been elicited by chance in any one of the five quarters of the individual waves, it makes more sense to ignore the carried-forward records and associate the individual records with waves not quarters—if it can be assumed that individuals whose carried-forward record is updated can be excluded from the immigrant/native comparison without significant bias. The year-quarter covariates (dummy variables for whether or not an 'observation' is in a particular year-quarter) would be replaced by 'wave covariates' that identify the wave of any individual record. This revision would also simplify the calculation of the non-standard 'standard errors' of Table 2—by not needing the recondite technique designed to handle the 100% autocorrelation of identical 'observations' on the same individual (Schmidheiny, 2013).

A remarkable absence of exploratory data analysis (EDA)

Until further notice, this paper will focus on Pubs 2 & 3, and will (in order to keep things relatively simple) ignore the 'observation' v. individual ambiguity. The sample size of 3.5 million will therefore be treated as if it were an ordinary collection of random samples (which it is not) of uncorrelated observations (which they are not) i.e. as if they were individual independent records in a conventional data-base whose message we would like to ascertain. In particular, the question of what are the explanatory variables that influence an individual's Yes or No answer to the "whether receiving or claiming state benefits" survey question?

It is remarkable that all the CReAM studies get under way without any preparatory EDA of the data. For instance, prior to the formulation of Model 1, there appears to have been no 4 x 44 tabulation of the cross-sectional frequencies that could reveal any year-to-year variation in the four frequencies for quarter q (a_q for immigrant Yes, b_q for native Yes, c_q for immigrant No, d_q for native No). The first EDA step would have been to plot the two time-series of Yes proportions $p_{Yqi} = a_q/(a_q + c_q)$ and $p_{Yqn} = b_q/(b_q + d_q)$ to see whether there is any trend in the differences $\Delta_q = p_{Yqi} - p_{Yqn}$. To the extent that LFS can be trusted, these differences can be regarded as unbiased quarterly estimates of the difference between the corresponding population proportions. If D&F had uncovered such a trend, their formulation of Model 1 might have been inhibited as a specification too rigid to represent the complex dynamics of the comparison.

The CReAM models constrain the definition of probability difference ('differential in probability')

If we describe the population proportions corresponding to p_{Yqi} and p_{Yqn} as 'probabilities', then Δ_q is an estimate of a 'differential in probability'. For the data of Pubs 2 & 3, there are 44 such differences—one for each quarter's cross-section of the LFS data. There is, however, only one 'differential in probability' in Pub 2's Model 1—a model that fits the values of y (the 1/0 dummy variable for Yes/No) with a putative probability $\beta I + \gamma_q$ (i.e. γ_q for a native in quarter q when I = 0, and $\beta + \gamma_q$ for an immigrant when I = 1; the parameter α in D&F's statement of Model 1 can be omitted if we use 44 dummies for the $\{\gamma_q\}$.) Table 1 exhibits the least-squares estimate of β as −17.8% . Model 1 is, in effect, a prior assertion of lack of interest in the values of Δ_q . This is how Pub2 acknowledges the rigidity of Model 1 in its incorporation of a single parameter β that plays the same role in every quarter:

> *When we regress our indicator variable only on immigrant status and time dummies, the coefficient indicates the percentage points difference in the probability of receiving benefits…between immigrants and natives observed* ***at the same moment in time****.* (p.10; my emphasis)

Stone (2013) found that the coefficient has an explicit expression as the weighted average $w_1\Delta_1 + \ldots + w_{44}\Delta_{44}$ (where weight w_q is proportional to the product $n_q p_q(i) p_q(n)$ of the number of observations n_q and the proportions $p_q(i)$, $p_q(n)$ of immigrant and native records in quarter q. In response, D&F (2014a) reworded the 'same moment in time' interpretation—the additivity assumption (the plus sign in $\beta + \gamma_q$) is not to be questioned but the consequent estimate of β is to be seen (not without ambiguity of description) as

> *the weighted averaged difference across quarters in the probability of receiving benefits … between immigrants and natives, conditioning on fluctuations in welfare receipt over time that affect immigrants and natives alike*.
> (D&F 2014a Appendix, and Pub 3 p.9)

The normality-probit model of Pub 3 also constrains the definition of probability difference—my Appendix 1 gives the rather technical details. However, despite the caveats associated with Model 1, there is little doubt that the large percentage differences 23.3%, 17.8% and 17.4% in Table 1 mean that there must be large cross-sectional differences in the 'observed' proportions of Yes responses of natives and immigrants, but we are not told whether these vary from quarter to quarter with implications for the dynamics of the comparison.

How well do the CReAM models fit the data?

As noted by Stone (2013), Pubs 1 and 2 do not address this question. For the least-squares fit of the Model 1 of Pubs 1 and 2 in Table 1, it could have been answered quite straightforwardly by further EDA—by seeing how close the estimates of γ_q for natives and $\beta + \gamma_q$ for immigrants were to the corresponding observed proportions p_{Yqn} and p_{Yqi}. But the assessment of fit is not so simple for Model 2, which D&F describe as 'counterfactual' and where the inclusion of personal covariates such as age and sex makes the comparison of immigrant and native more 'like with like'. In Model 2, the single term γ_q of Model 1 representing the year-quarter covariates is replaced by a general linear expression $\boldsymbol{\gamma}^T \mathbf{x}$ that includes the additional contributions of the personal covariates (components of vector **x).**

For such models, econometric studies routinely play with different selections of covariates and let the traditional statistical measure R^2 of overall model performance guide the final choice—which is why Stone (2013) commented on the absence of any R^2 values. (In a least-squares fit of a linear model with a constant as one of the explanatory variables, R^2 is the square of the correlation coefficient between the dependent variable and its fitted value, but it can also be defined by equating $1 - R^2$ and the ratio of the sum of the squares of residuals to the analogous sum for the model that simply fits a constant to all the observations). Why did D&F not reveal the missing values of R^2 (standard output of any least-squares software) but went so far in as to reject their relevance to their modelling? :

> *When we condition on observables* [the additional covariates], *what matters is not R^2 per se, but how different characteristics may affect benefit take-up and whether these characteristics are correlated with immigrant status.* (D&F, 2014a, Appendix)

The following argument shows that the figures that were given are enough to bridge the gap of not being given the relevant R^2. To see this, consider both Pub 2's Model 2 for the estimate −8.4% and the nested sub-model, call it Model 2/0, that omits the immigrant/native status variable (equivalent to putting $\beta = 0$ in Model 2). If the generic residual r_0 of Model 2/0 changes to $r_0(1-\varphi)$ in Model 2, we can say that Model 2 is, for that observation, closer by

φr_0 to the 1 or 0 value of y (if r_0 were to increase, φ would be negative). Appendix 2 shows firstly that the (necessarily non-negative) fractional reduction of the sum of squares of residuals is equal to the weighted average of φ with weights proportional to r_0^2. Appendix 2 then uses Table 1 data to show that the fractional reduction (and hence the more easily interpretable formulation as the weighted average of φ) is almost certainly less than 1%. In other words, the immigrant/native status variable accounts for only a small fraction of the otherwise unexplained variation. The uncertainty will only be resolved when CReAM divulges the values of R^2 for Model 2 and Model 2/0.

Can we 'relax' whenever we add explanatory variables to an empirical model?

Much of the following quotation from D&F (2014a) could be questioned at length for its explicitly subjective statements. There is, however, an implicit assumption running through it. It is one that merits exposure by concrete counter-example rather than disciplinary generality. That is because relaxation is tacitly and too often countenanced in econometric studies that use what can be called 'accountancy models'—additive linear models in which each element of a string of terms is there to 'take account of' the influence of a particular factor (explanatory variable). There is a conjunction of computational convenience (least-squares software takes care of the computation) and economy of thought about what is being done when an extra variable is introduced and tested, by rerunning the software, to see whether it improves the explanatory performance (e.g. R^2) of the fitted formula by 'taking due account of' the factor—in which case it is included and its influence is 'captured' (but is otherwise excluded without consequence). As the quotation suggests, that has become a 'standard procedure':

> *To capture differences between immigrants and natives in demographic characteristics, we condition on gender and a quadratic in age. Again, this is a standard procedure. Of course, it implies an assumption about functional form—which we believe is not implausible but at the same time simple and transparent. One could relax functional form assumptions by including a full set of dummy variables for age, and interact them with gender dummies, or use matching type estimators. Using such estimators, results show an even larger difference in welfare and transfer receipt between immigrants and natives than reported in our Table 3* [in Pub 2]. *For instance the gap resulting from a fully interacted model specification is −0.125 for immigrants arriving since 2000, compared to the estimates of our more restricted specification reported in the Table, which gives an estimate of −0.084* (D&F, 2014a, Appendix)

By implicitly welcoming the *even larger difference −0.125,* D&F appear to suggest that adding more variables (by 'relaxing' the 'functional form') will always get you nearer some more plausible 'truth'. If so, they are mistaken, as the following counter-example (an appropriate proxy for the CReAM results) demonstrates. Model B is the proxy for Model 2 in Pub 2 with year-quarters, a quadratic in age & sex as covariates, whereas Model C is the proxy for the relaxation that conditions on the extra covariates in 'age', sex and their interaction. Here is the counter-example, whose implication extends well beyond the present context:

> Suppose the 'true' model T is $Y = \beta X_B + \gamma X_C + \delta X_T + \varepsilon$. Model B uses only variable X_B, Model C uses only X_B and X_C, but model T uses all three variables. If the least-squares estimating equations (with standardized variables and a sample size of millions) are $\mathbf{R\theta} = \mathbf{r}$ where $\mathbf{r} = (-0.084, 0.695, -0.414)^T$ and $\mathbf{R}$ is the 3 x 3 correlation matrix

	X_B	X_C	X_T
X_B	1.000	0.100	−0.001
X_C	0.100	1.000	−0.682
X_T	−0.001	−0.682	1.000

> the successive estimates of β in the progression $B \rightarrow C \rightarrow T$ are −0.084, −0.155 and 0.100.

The relaxed Model C's −0.155 is further away from the true value 0.100 than Model B's −0.084. The only gap that always narrows when more explanatory variables are added to a linear model is the gap between observed and fitted values measured by the residuals measure $1 - R^2$.

Are the cross-sectional differences $\{\Delta_q\}$ reflected in the net fiscal balance differences?

Pub 2 made no theoretical case for the assumption of additivity in Model 1, which may be why D&F (2014a) was happy to note (p.22) that the estimate of the immigrant/native differential was indeed a weighted average of the observable and interpretable differences $\{\Delta_q\}$ in the 44 quarters. The weights $\{w_q\}$ are proportional to $n_q(i)p_q(n)$

where $n_q(i)$ is the number of immigrants in quarter q. Table 1a of Pub2 suggests that $n_q(i)p_q(n)$ increased by a factor of about 9 between 2001 and 2011, while Figure 1 here suggests that Δ_q may have become increasingly negative over the same period, particularly after the 2007/08 financial crash. If so, the larger weights would be going to the most negative values of Δ_q and the resulting estimate of β would be decidedly negative—in line with the CReAM estimates of −17.8% and −17.4%.

The possibility that, contrary to such speculation, Model 1 with its constant probability difference might still be a fair and inferentially useful description of the data can only be determined when CReAM publishes the 88 cell proportions of which the 44 values of Δ_q are the differences. Meanwhile, its relationship to the spectacular jump in Figure 1 after the 2007-08 can be a legitimate matter of conjecture by commentators such as Robert Rowthorn. My admittedly-uninformed conjecture is that the jump in fiscal balance difference may be a reflection of some immigrant/native interplay over time that was overlooked in the CReAM studies. Such an 'interaction' would raise questions that cannot be appropriately addressed by time-indifferent econometric models.

Figure 1. **The *per capita* 'immigrant minus native' differences from Table 4.**

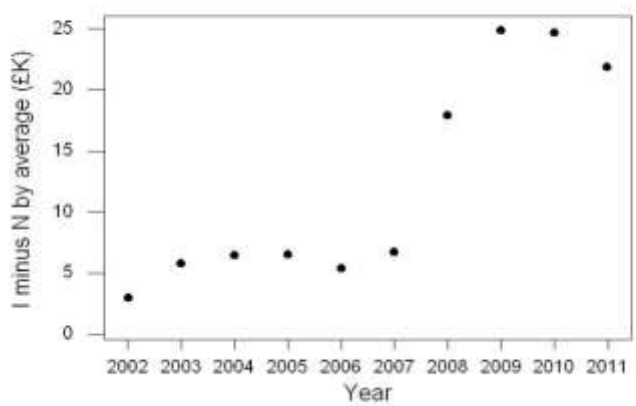

Table 4. ***Per capita*** **average-measure fiscal balance, £ at constant 2011 prices.**

Year	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Post-2000 immigrants	276	252	381	263	686	967	1221	708	5	−120
Natives	−2658	−5554	−6047	−6274	−4665	−5746	−16687	−24224	−24696	−22028

Refusing to recognise an elephant in a room

The reluctance of the environment panel in the attic of the Bush Theatre to grasp or concede the significance of population numbers was outclassed by the hostility of the audience assembled for the BBC's opposition party Election Debate. The hostility was directed at the party leader who dared to mention the population factor for pressures on public resources and infrastructure. It was as if the studio air was imbued with an aggressive spirit of good feeling towards all immigrants—and as if the man was daring to refer to an elephant in the room that the audience was unwilling to recognize.

There is quite a herd of similarly intrusive elephants. For the Civitas health briefing (Stone, 2014), the elephant was in the PAC committee room when all parties refused to consider the near certainty (see Stone, 2015) that hundreds of billions may have been misallocated over decades by a grossly defective funding formula. For the present note, there is a smaller elephant in the space of public discourse about the CReAM studies—but one that is big enough, if unchallenged, to squeeze out the healthy controversy that should have been generated by those studies. The purpose of this note has been to maintain and, if possible, prolong the challenge.

D&F's response to the Civitas press release purported to find broad agreement between my critique and the non-econometric component of D&F (2013)—the estimation of net fiscal balance. It relegated my criticisms of its econometrics to an appendix and dealt with them there by polite rebuttal. In fact, Stone (2013) was severely critical

of the inadequately informative presentation of the fiscal balance tabulations and their evasion of per capita evaluations (now made obvious by seeing how things are calculated and tabulated in Rowthorn's book).

Conclusions

This note has shown that there is not much of 'what to believe' in the CReAM econometrics. There is, rather, too much of 'what's not to believe':

● The CReAM studies eschew exploratory data analysis that could reveal (who knows?) the need for a substantially different 'dynamic' description of the immigrant/native comparison.

● None of the CREAM studies justifies the grossly-multiple use of identical individual records from the LFS data-base (creating millions of 'observations') or demonstrates that this increase in 'sample size' does not result in artefactual bias in any immigrant/native comparison.

● There is no justification for readers to believe in the validity of the empirical (theoretically vacuous) models for 'like-with-like' comparison of immigrants and natives. (The onus of proof that trust in any chosen model is deserved lies with the study authors rather than readers.)

● The 'differential probability' estimates for the like-with-like comparisons vary appreciably from study to study with different choices of the explanatory covariates—probably a manifestation of the intrinsic lack of robustness of empirical model choice. (The small standard errors and the stars of statistical significance they generate express only one fact about these estimates—that they are the predictable consequences of fitting any mildly plausible model to the supposed millions of 'observations'.)

● The CReAM studies do not state the conventional R^2 measure of model performance, but they do give enough information to establish that, for the like-with-like comparisons, there is a great deal of room (in the space of unexplained variation) for different empirical models to give very different estimates.

● Reliable estimation of immigrant/native differences are probably unobtainable without statistically-principled application of small-scale data collection—enough to establish the fiscal activity of random samples of immigrants and natives without reliance on self-reported LFS data. (Such an effort would be well beyond the resources of this commentator).

Appendix 1

The 'linear probability model' of Pubs1 & 2 can generate fitted values of 'probability' unrealistically outside the range 0 to1 when fitted by least-squares. In Pub 3, it was replaced by a 'normality-probit model', in which Prob(y =1) = $\Phi(\beta^*I + \boldsymbol{\gamma}^{*T}\mathbf{x})$—the cumulative probability in the interval $(-\infty, \beta^*I + \boldsymbol{\gamma}^{*T}\mathbf{x})$ of a standard normal random variable. (Φ is the 'cumulative distribution function' and I is the dummy variable for immigrant/native status). Pub4 does not say how the parameters β^*, $\boldsymbol{\gamma}^*$ were estimated i.e. whether by maximum likelihood or least-squares estimation of the non-linear function Φ. According to Pub4, estimates of the immigrant/native probability differential were obtained as *marginal effects from [the] probit model, computed at the mean value of all the other regressors (*p.23). I interpret this to mean that they were obtained by inserting the estimates of β^* and $\boldsymbol{\gamma}^*$ in the partial derivative $\partial\Phi(\beta^*z + \boldsymbol{\gamma}^{*T}\mathbf{x})/\partial z$—evaluated at $z = I^{ave}$ and $\mathbf{x} = \mathbf{x}^{ave}$ where 'ave' denotes sample average. Apart from the change of probability model, there are also the 'relaxation' changes to the Model 2 covariates **x** of Pub2 to embrace more conditioning covariates:

> *'Age' is a set of dummy variables for 5 years age groups starting at age 16, 'Gender' are gender dummies interacted with the full set of age dummies.* (Pub 3, p.23)

Note that, with a quadratic for age, Pub2's Model 2 is probably practically equivalent (as far as 3-significant-figure estimation is concerned) to a sub-model of the one with 'age' (given such narrow age-bands). This equivalence is assumed as an element in the 'relaxation' counter-example. Table 3, showing how well $\Phi(z)$ is approximated by $0.5 + z/(2\pi)^{½}$, indicates that the change from linear probability models to probit models will have little effect on the estimates if it can be supposed that most of the fitted probabilities are in the interval (0.20, 0.80):

Table 3. Best linear approximation of probability Φ

$\Phi(z)$	0.10	0.15	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.85	0.90
$0.5 + z/(2\pi)^{½}$	−0.01	0.09	0.16	0.29	0.40	0.50	0.60	0.71	0.84	0.91	1.01

This approximation may explain the near equality of the Model 1 estimates −17.8% and −17.4% in Table 1. It may, however, put the onus of explaining the near doubling of the Model 2 estimates (from −8.4% to −15.5%) on either some change in the data-base (dropping 44,000 observations out of 3.5 million?) or on the 'relaxation' of the model already implicated in the counter-example (by replacing the quadratic in age by 5-year age-bands for each sex).

Appendix 2

The fractional reduction of the sum of squares of the residuals of Model 2/0 is $1 - \text{Sum}(r_0 - r_0\varphi)^2/\text{Sum } r_0^2$ which can be seen to be the weighted average of $2\varphi - \varphi^2$ with weights proportional to r_0^2. The geometry of least-squares estimation of Model 2 implies that the vector with components $r = r_0 - r_0\varphi$ is perpendicular to the vector with components $r_0 - r = r_0\varphi$, whence $\text{Sum } r_0\varphi(r_0 - r_0\varphi) = 0$ or $\text{Sum } r_0^2(\varphi - \varphi^2) = 0$, proving the equality of the fractional reduction and the weighted average of φ itself. If the standard error 0.1% for the −8.4% estimate is a rounded value, the t-value is less than 8.4/0.05 = 168. If it can also be assumed that the 0.1% is not less than it would have been without the Schmidheiny correction for autocorrelation, the ordinary-least-squares t-value is also less than 168. Galbraith and Stone (2011) showed that, for least-squares estimation, the fractional reduction of the sum of squares of residuals by the addition of one more parameter (here the β) to a model with n observations and p parameters is $t^2/(t^2 + n - p - 1)$. It follows that, with n = 3,495,478 and p = 47, the fraction is less than 1%.

Acknowledgements

I am grateful to Professor Bob Rowthorn (Department of Economics, University of Cambridge) for the estimates in Table 4, which come from Figures 5.1c & 5.1d of Rowthorn (2014b). Dr Rex Galbraith (UCL) kindly did the calculations for the counter-example, and Professor David Coleman (University of Oxford) gave sound advice.

References

Browne, L. and Alstrup, P. (2006) *What exactly is the Labour Force Survey?* http://webarchive.nationalarchives.gov.uk/20140721132900/http://www.statistics.gov.uk/downloads/theme_labour/What_exactly_is_LFS1.pdf ?

Civitas (2013) *"Schoolboy errors" in UCL report claiming fiscal benefit to immigration*. http://www.civitas.org.uk/press/PRimmigration.html

Dustmann, C., Frattini, T. and Halls, C. (2010) *Assessing the Fiscal Costs and Benefits of A8 Migration to the UK, Fiscal Studies*, **31**, 1-41. www.ucl.ac.uk/~uctpb21/Cpapers/DustmannFrattiniHalls2010.pdf .

D&F (2013) *The Fiscal Effects of Immigration to the UK*. London: Centre for Research and Analysis of Migration. www.cream-migration.org/publ_uploads/CDP_22_13.pdf .

D &F (2014a) *There are no "Schoolboy Errors" in our publication*. London: Centre for Research and Analysis of Migration. http://creamcomments.blogspot.co.uk/2014/01/there-are-no-schoolboy-errors-in-our.html .

D&F (2014b) *The Fiscal Effects of Immigration to the UK. The Economic Journal*, **123**, 1 - 51. Doi: 10.1111/ecoj.12181©

Galbraith, J. and Stone M. (2011) *The abuse of regression in the National Health Service allocation formulas: response to the Department of Health's 2007 'resource allocation research paper' (with Discussion)*. *J. R. Statist. Soc.* A, **174**, 517-528.

Rowthorn, R. (2014a) *A note on Professor Stone's comment on Dustmann and Frattini.* www.civitas.org.uk/pdf/AnoteonStone.pdf.

Rowthorn, R. (2014b) *Large-scale Immigration: Its economic and demographic consequences for the UK'*. www.civitas.org.uk/pdf/LargescaleImmigration.pdf .

Schmidheiny, Kurt. (2013) *Clustering in the Linear Model*. kurt.schmidheiny.name/teaching/clustering2up.pdf.

Stone, M. (2013) *Plain assumptions and unexplained wizardry called in aid of "The Fiscal Effects of Immigration to the UK''*. London: Civitas Online. www.civitas.org.uk/pdf/assumptionsandwizardry.pdf.

Stone, M. (2014) *National Auditing beats Parliamentary Accounting*. www.civitas.org.uk/pdf/CivOctRev5.

Stone, M. (2015) *The remarkable insignificance of NHS England's CCG-funding formula. Public Money and Management* (to appear in July or September).