# Analysis of Statistical Disclosure Control in Census 2001 Origin-Destination Tables

Natalie Shlomo
SDC Centre of the Methodology Group, ONS

## 1. Introduction

Origin-destination migration and work place tables were released as census 2001 outputs at the OA level of geography. Before their release, the tables underwent a perturbation scheme for statistical disclosure control in order to mask the small cell values of the table. According to the perturbation scheme, the cells with small values were adjusted independently upwards and downwards based on prescribed probabilities. The totals and sub-totals in the table (i.e., margins) were obtained by summing across cells after the perturbation of the small cells. The perturbations have an expected mean of zero and a variance that is proportional to the number of small cells that were adjusted. The probability mechanism that was devised for carrying out the perturbations was defined in a Random Number Table. The table contains a relatively small number of $N$ records and each record contains a random uniform number between 1 and $N$ that was generated by an SQL routine. The small cell values are perturbed upwards (downwards) if the random number is greater (smaller) than $N$ times the prescribed probability. This is equivalent to drawing a random number between 0 and 1 with perturbed values going up (down) if the prescribed probability is smaller (greater) than the random number. Since the Random Number Table has a relatively small and finite number of records containing the random uniform numbers, all perturbations of the origin-destination tables are carried out on one realisation of the stochastic process for generating random uniform numbers and therefore biases may be present.

For a table of origin OA's to destination OA's undergoing the perturbation scheme, an initial starting entry for the Random Number Table is generated for the perturbation of the first cell. All further small cell values are perturbed upwards or downwards according to the consecutive random uniform numbers in the Random

Number Table following the initial starting entry. For practical reasons, the initial starting entry was determined by a function of the digits and letters in the code of the LAD to which the origin OA belongs and the code of the LAD to which the destination OA belongs. This was to ensure that the same perturbation of an origin-destination table is carried out for any running of the process. Note that the only source of randomness in this perturbation scheme is the Random Number Table (which is itself one realisation of a stochastic process) since the initial starting entry depends on the codes of the LAD's.

Recently, the census 2001 origin-destination migration and work place tables were released at the ward geographical level. Because of the perturbation scheme, the aggregated totals summed up from the OA tables are expected to be different from the totals obtained at the ward level. In general, the same perturbation scheme is carried out at higher geographical levels but since there are fewer small cells, this leads to fewer perturbations and more exact totals. Therefore, there is a need to establish the general degree to which the totals differ in order to inform users of the impact of the perturbation scheme on the origin-destination tables. In addition, users are advised to use the minimum number of cells when aggregating combinations of geographical areas or when calculating sub-groups of the population (see user guidelines released with the origin-destination tables at the ward level).

Upon release of the tables at the ward level, the SDC Centre of the Methodology Group was asked to prepare guidelines for users in order to estimate the extent of the differences from the raw totals when aggregating from the perturbed OA origin-destination tables. In addition, the SDC Centre and the Census Outputs Division examined a few tables for possible biases that may have been introduced into the perturbation scheme which would lead to larger differences than expected. In particular, there are two possible sources which may have biased the perturbation scheme by introducing dependencies into the process:

- It was found that the distribution of the initial starting entries based on a function of the codes for the combination of the origin and destination LAD's is not uniform and as a result most of the initial starting entries for the perturbation of the tables were carried out within a narrow band of the Random Number Table.

- Because of the nature of the process for generating random uniform numbers for the Random Number Table, there may be discrepancies between the expected prescribed probabilities and the actual probabilities obtained from the one realisation of the stochastic process.

Section 2 describes the data that were used for analysis. Section 3 details the statistical testing with some results carried out on ward migration origin-destination tables for possible biases that may have been introduced. Section 4 explains the proposed method for assessing the expected level of differences from the raw totals when aggregating perturbed origin-destination tables. Recommendations will be presented in Section 5.

## 2. Data Used

The main tables that were analysed for this work were migration tables MG201 for two origin LAD's: 00AB and 00BL. The tables include sub-populations according to sex-age distributions. Both the original and perturbed tables were provided as well as aggregated forms of the tables in an Excel spreadsheet. The number of wards having a non-zero number of persons migrating to other wards is 2,844 for table MG201_00AB and 3,434 for table MG201_00BL. Combining the two tables, we obtain 6,278 wards with persons migrating to other wards. For each ward to ward combination, an age-sex distribution is given containing 32 cells, thus a total of 200,896 cells in the table, thereof only 21,268 cells with a non-zero entry.

In addition, the workplace table W206 was provided containing seven LAD's in the West Midlands. The table has 12 columns according to employment status and sex. The number of origin wards to destination wards in all of the LAD's having a non-zero number of workers migrating for work is 31,447, thus there is a total of 377,364 cells in the table, thereof only 58,475 cells with a non-zero entry.

For the purpose of assessing the extent of the differences from the raw totals caused by the perturbation scheme, we generated from the tables a large number of groupings of origin wards starting from 5 wards to about 600 wards. The groupings were formed by aggregating neighbouring origin wards since this would be consistent with the usage of the table by the users. Note that the groupings are dependent since each origin ward may appear in many groupings. For the origin LAD of 00AB, 126,907 groupings were formed, for the origin LAD of 00BL, 125,549 groupings were formed, and for W206 of West Midlands with seven LAD's, 691,396 groupings were formed.

### 3. Checking for Possible Biases in the Perturbation Scheme

A number of statistical tests were carried out to check for possible biases in the tables:

- Based on the 21,268 non-zero cells in the combined origin-destination migration MG201 ward table, 83.4% of the cells were perturbed. For these cells, we test to see if the mean of the perturbations is zero using PROC UNIVARIATE in SAS. The procedure tests the null hypothesis of a mean equal to zero based on a t-test. The mean perturbation is 0.0230 with a standard error of 0.0089, and the t-statistic is 2.576 which is rejected with a p-value of 0.010. For MG201_00BL, the mean perturbation was .0097 with a standard error of .0117 and therefore the null hypothesis was accepted and for MG201_00AB, the mean perturbation was 0.0395 with a standard error of 0.0138 and again the null hypothesis was rejected with a p-value of .004.

- Using the same 21,268 non-zero cells in the combined migration MG201 origin-destination ward table, we checked to see if the perturbations were carried out under the correct prescribed probabilities. In general, there are two prescribed probabilities in the scheme. An exact binomial proportional test was performed in SAS using the PROC FREQ procedure where the null hypothesis is: $p = p_0$ and the p-value is calculated using the Binomial CDF. On the combined origin-destination migration ward table, for the first prescribed probability, the null hypothesis was rejected with a p-value equal to 0.001 and the proportion of cells that were actually perturbed was significantly different than the prescribed probability. For the second prescribed probability the null hypothesis was accepted. The same results are obtained for MG201_00AB with a p-value of

0.011 for rejecting the null hypothesis on first prescribed probability. For MG201_00BL, both of the tests on the prescribed probabilities were not significant and the null hypothesis was accepted.

- The expected difference between the raw and perturbed totals for a ward $i$, $e_i = T_i^R - T_i^P$, should be zero for an unbiased perturbation scheme. In order to test whether the difference is significant, a confidence interval was calculated for each $e_i$ based on the variance and assuming that the perturbation scheme can be approximated by the normal distribution: $CI_i = 1.96\sqrt{\text{var}(e_i)}$. If the confidence interval contains zero, then it can be said that there is no significant difference between the raw and perturbed total in the ward. If, however, the confidence interval does not contain zero, then it can be said at a 95% significance level that there is a difference between the raw and perturbed total for the ward and the difference falls between $e_i - CI_i$ and $e_i + CI_i$.

Of the 6,278 origin wards in the combined migration origin-destination table, only 34 (0.54%) origin wards had no perturbation carried out. The differences between the raw and perturbed totals were significantly different from zero for 188 (3.0%) origin wards. When grouping the origin wards in neighbouring groups of 5, 53 (4.2%) out of 1,255 groupings had differences that were significantly different from zero, and for neighbouring groups of 10, 31 (4.9%) out of 628 groupings had differences that were significantly different from zero. Since we are constructing 95% confidence intervals, none of these tests suggest biases in the tables. Note that the average number of non-zero cells for an origin ward is 3.4 (21,268 non-zero cells in 6,278 origin wards). Therefore, the difference between the raw and perturbed totals for one ward will typically lie within the confidence interval and we do not see a strong case for biases. As we aggregate across the origin wards, more discrepancies between the raw and perturbed totals may occur outside of the confidence intervals.

Table 1 presents the total number of non-zero cells, the number of cells that were perturbed and the raw and perturbed totals for the combined LAD's and each

LAD separately for the MG201 migration table. At this LAD level, the differences between the raw and perturbed total that were significant are shaded. The actual percent difference between the raw and perturbed totals for MG201_00AB was 2.3% and for MG201_00BL 0.4%. The differences can potentially become larger for sub-groups of the table and as the geographical areas get smaller. For the combined tables the percent difference is 1.1%.

**Table 1: Summary of Cell Counts and Totals for Migration Origin-Destination Table MG201 – LAD 00AB and LAD 00BL**

| LAD | | Total | Males | Females |
|---|---|---|---|---|
| Total | Num. of non-zero Cells | 21,268 | 10,415 | 10,853 |
| | Cells Perturbed | 17,743 | 8,718 | 9,025 |
| | % Cells Perturbed | 83.4 | 83.7 | 83.2 |
| | Raw Total | 43,805 | 21,554 | 22,251 |
| | Perturbed Total | 44,294 | 21,553 | 22,741 |
| | Difference | -489 | 1 | -490 |
| | % Difference | -1.12 | 0 | -2.20 |
| 00AB | Num. of non-zero Cells | 9,467 | 4,608 | 4,859 |
| | Cells Perturbed | 8,225 | 4,025 | 4,200 |
| | % Cells Perturbed | 86.9 | 87.3 | 86.4 |
| | Raw Total | 16,241 | 7,858 | 8,383 |
| | Perturbed Total | 16,615 | 7,902 | 8,713 |
| | Difference | -374 | -44 | -330 |
| | % Difference | -2.30 | -0.56 | -3.94 |
| 00BL | Num. of non-zero Cells | 11,801 | 5,807 | 5,994 |
| | Cells Perturbed | 9,518 | 4,693 | 4,825 |
| | % Cells Perturbed | 80.7 | 80.8 | 80.5 |
| | Raw Total | 27,564 | 13,696 | 13,868 |
| | Perturbed Total | 27,679 | 13,651 | 14,028 |
| | Difference | -115 | 45 | -160 |
| | % Difference | 0.42 | 0.33 | 1.15 |

Table 2 presents the total number of non-zero cells, the number of cells that were perturbed and the raw and perturbed totals for the work place W206 table containing seven LAD's. The differences between the raw and perturbed total that were significant are shaded.

**Table 2: Summary of Cell Counts and Totals for Workplace Origin-Destination Table W206 – Seven LAD's**

| W206 | | Total | 00GA | 00GF | 39UB | 39UC | 39UD | 39UE | 39UF |
|---|---|---|---|---|---|---|---|---|---|
| Totals | Num. of non-zero Cells | 58,475 | 13,012 | 13,422 | 7,182 | 7,129 | 3,895 | 8,996 | 4,839 |
| | Cells Perturbed | 43,492 | 9,616 | 9,663 | 5,597 | 5,507 | 2,832 | 6,446 | 3,831 |
| | % Cells Perturbed | 74.4 | 73.9 | 72.0 | 77.9 | 77.2 | 72.7 | 71.7 | 79.2 |
| | Raw Total | 291,105 | 81,289 | 74,380 | 27,026 | 26,833 | 17,219 | 46,146 | 18,212 |
| | Perturbed Total | 291,438 | 81,165 | 74,074 | 27,272 | 27,009 | 17,300 | 46,333 | 18,285 |
| | Difference | -333 | 124 | 306 | -246 | -176 | -81 | -187 | -73 |
| | % Difference | -0.11 | 0.15 | 0.41 | -0.91 | -0.66 | -0.47 | -0.41 | -0.40 |
| Males | Num. of non-zero Cells | 32,995 | 7,505 | 7,602 | 3,958 | 3,929 | 2,171 | 5,157 | 2,673 |
| | Cells Perturbed | 25,173 | 5,726 | 5,661 | 3,124 | 3,075 | 1,606 | 3,820 | 2,161 |
| | % Cells Perturbed | 76.3 | 76.3 | 74.5 | 78.9 | 78.3 | 74.0 | 74.1 | 80.8 |
| | Raw Total | 160,884 | 44,698 | 40,868 | 15,760 | 15,256 | 9,300 | 24,896 | 10,106 |
| | Perturbed Total | 160,315 | 44,394 | 40,576 | 15,799 | 15,290 | 9,349 | 24,815 | 10,092 |
| | Difference | 569 | 304 | 292 | -39 | -34 | -49 | 81 | 14 |
| | % Difference | 0.35 | 0.68 | 0.71 | -0.25 | -0.22 | -0.53 | 0.33 | 0.14 |
| Females | Num. of non-zero Cells | 25,480 | 5,507 | 5,820 | 3,224 | 3,200 | 1,724 | 3,839 | 2,166 |
| | Cells Perturbed | 18,319 | 3,890 | 4,002 | 2,473 | 2,432 | 1,226 | 2,626 | 1,670 |
| | % Cells Perturbed | 71.9 | 70.6 | 68.8 | 76.7 | 76.0 | 71.1 | 68.4 | 77.1 |
| | Raw Total | 130,221 | 36,591 | 33,512 | 11,266 | 11,577 | 7,919 | 21,250 | 8,106 |
| | Perturbed Total | 131,123 | 36,771 | 33,498 | 11,473 | 11,719 | 7,951 | 21,518 | 8,193 |
| | Difference | -902 | -180 | 14 | -207 | -142 | -32 | -268 | -87 |
| | % Difference | -0.69 | -0.49 | 0.04 | -1.84 | -1.23 | -0.40 | -1.26 | -1.07 |

In summary of the above analysis on the tables that were examined, there appears to have been very slight biases introduced into the perturbation scheme. In particular, the second test showed that some of the perturbations did not follow the prescribed probability scheme. Although the biases are not severe, for future usage of this perturbation scheme the Random Number Table that is used to define the perturbations should be examined for large discrepancies from the expected prescribed probabilities and a more stochastic process should be introduced for generating the random uniform numbers. In addition, the initial starting entry for the Random Number Table needs to be redefined to ensure a uniform distribution so that all records of the Random Number Table have equal probability of being an initial starting entry. One suggestion is to define for each combination of origin LAD to destination LAD a random uniform number $u$ which will be placed in a special look-up file and saved for all processing of the tables. This results in a large file. Another alternative is to generate a random uniform number $u$ for each LAD separately and define $z = u^O + u^D$ where $u^O$ is the random uniform number for the origin LAD and $u^D$ is the random uniform number for the destination LAD. The starting value $\tilde{z}$ will equal $z$ if $0 \le z \le 1$ and $z - 1$ if $1 < z \le 2$. This method results in the same initial starting entry for an origin LAD to a destination LAD when it is the other way around.

## 4. Assessing the Errors Caused by Perturbations

Error is defined as the difference between the raw total and the perturbed total obtained when aggregating cells from a perturbed Origin-Destination table. To provide user guidelines for assessing the extent of the errors , care should be taken not to divulge the perturbation scheme. In addition, the biases that may have been introduced into the scheme means that theoretical confidence intervals may not give the correct coverage for the errors. Therefore, a more heuristic approach was taken to provide users with guidance for assessing the errors caused by the perturbations. The proposed method is based on a functional relationship between the perturbed totals aggregated from ward Origin-Destination tables and the $5^{th}$ and $95^{th}$ percentiles of the error distribution. This will allow users to calculate a "perturbation interval" around each perturbed total aggregated from the ward Origin-Destination table which will cover approximately 90% of the expected differences from the raw totals (i.e., errors).

For this analysis, we used the tables examined in Section 3 after generating groupings $j$ of the wards (see section 2). These groupings were sorted in ascending order according to the perturbed total, $T_j^P$. Perturbed totals less than 20 were initially not included in the analysis since it was assumed that these totals would need to undergo a separate analysis, but the final models include this grouping. The groupings $j$ in each of the tables were then formed into bands denoted by $k$. Each band had about 2,000 groupings for the totals and calculating the 5th and 95th percentiles of the error distribution.

In each band $k$ we calculated:

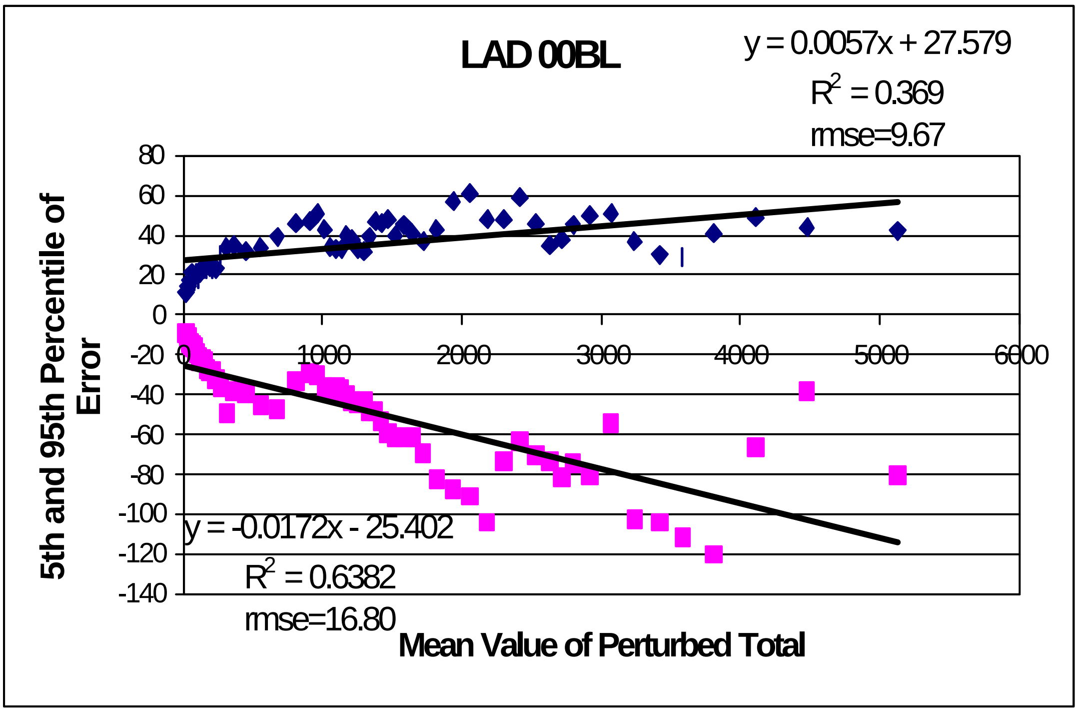
- the mean value of the perturbed total, $\overline{T}_k^p = \frac{1}{n_k}\sum_{j\in k} T_j^P$ where $n_k$ is the number of groupings in band $k$.

- the 5th and 95th percentile of the errors, $e_j = T_j^R - T_j^P$ for all $j \in k$.

Two separate regression models were fitted for the 5th percentile and the 95th percentile of the error distribution on the mean value of the perturbed totals $\overline{T}_k^p$ in band $k$. We used as the explanatory (independent) variable the mean value of the perturbed total since this is the information that is available to the users. The lower bound for the perturbation interval is obtained by fitting a linear regression where the dependent variable is the 5th percentile of errors $e_j$ in band $k$. The upper bound for the perturbation interval is obtained by fitting a linear regression where the dependent variable is the 95th percentile of errors $e_j$ in band $k$. To improve the fit of the regression models with respect to their $R^2$ and mean squared error of the predicted values, transformations on the variables were carried out. Since the variance of the perturbations is proportional to the number of perturbed cells which is approximately proportional to the perturbed total, we fit a linear regression model on the square root of the mean value of the perturbed total. In addition, we examined a multiplicative regression model by transforming both the error percentiles and the mean value of the perturbed total to the log which is more general and includes the square root as a special case.

In Figures 1 to 3, examples of some of the regression models that were fitted for the migration tables MG201_00BL and MG201_00AB are shown. The results of the regression analysis are summed up in the tables presented in the Appendix and an analysis of the residuals is shown for the pooled regression model after Figure 4.

**Figure 1 : Regression of Percentiles of Errors on Mean Value of Perturbed Totals – Migration Tables MG201_00BL and MG201_00AB**



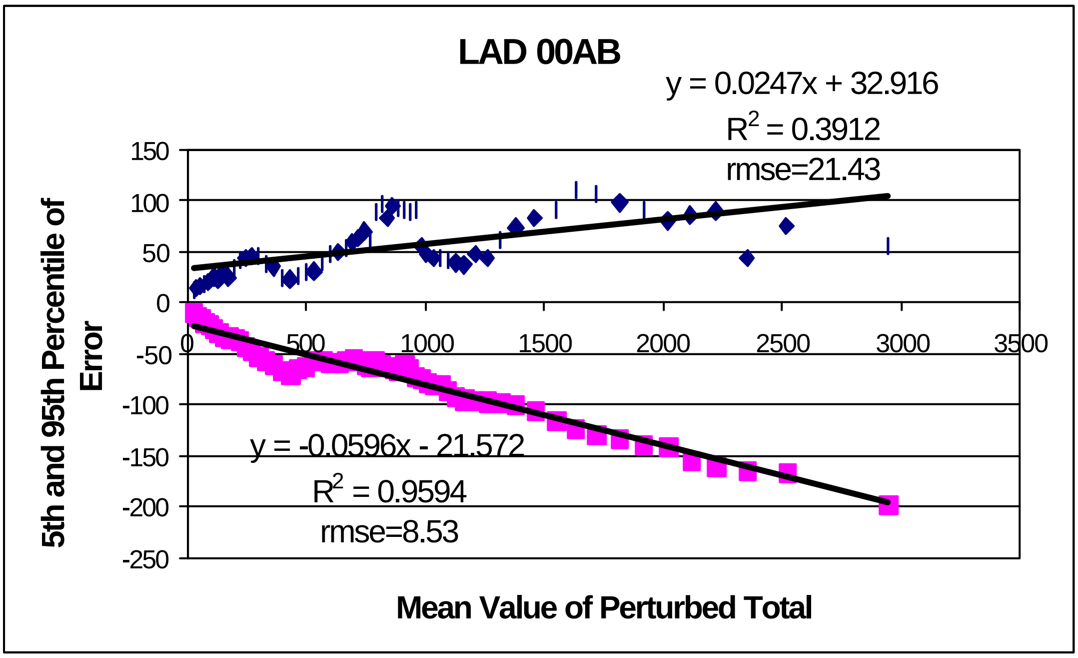
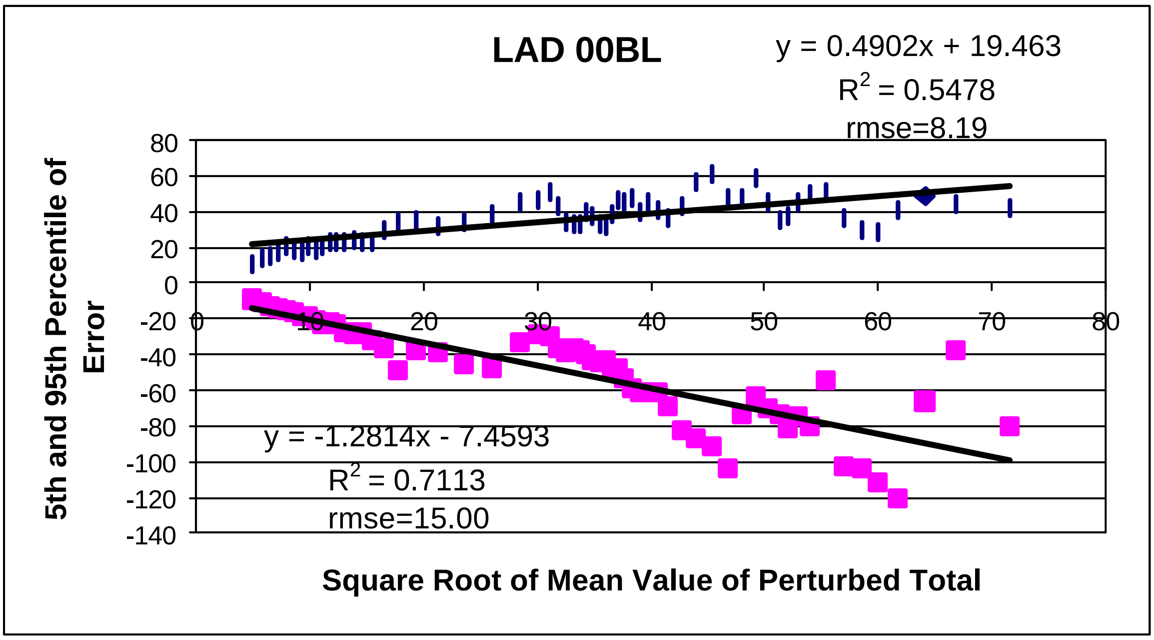



**Figure 2: Regression of Percentiles of Errors on Square Root of Mean Value of Perturbed Totals – Migration Tables MG201_00BL and MG201_00AB**



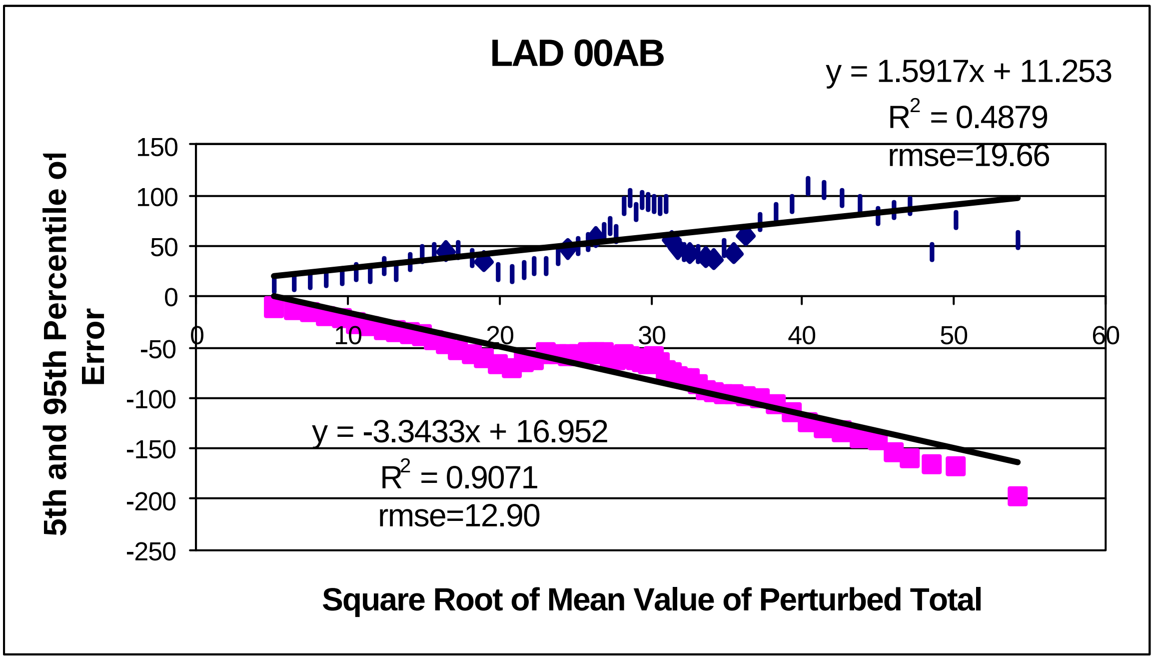
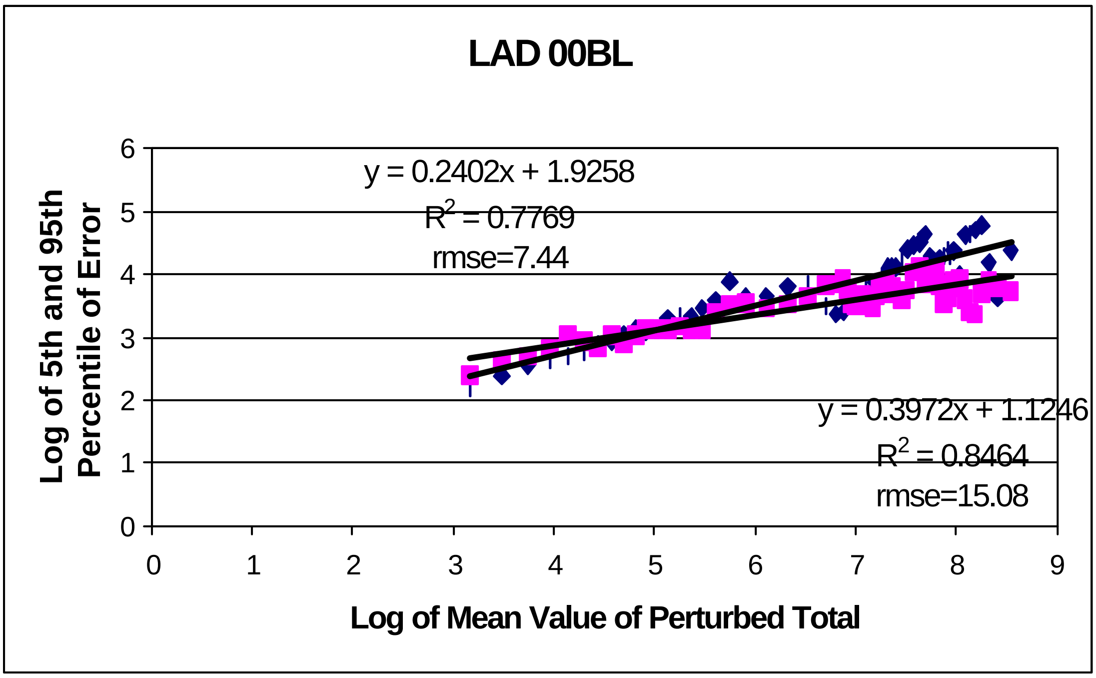
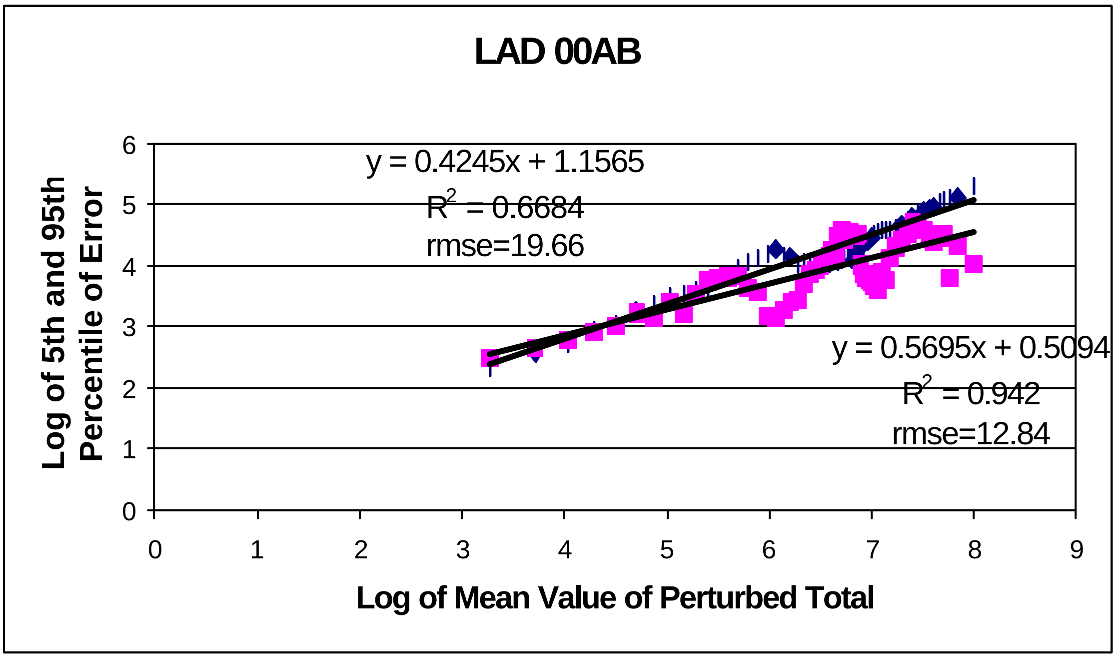



**Figure 3: Regression of Log of Percentiles of Errors on Log of Mean Value of Perturbed Totals– Migration Tables MG201_00BL and MG201_00AB**



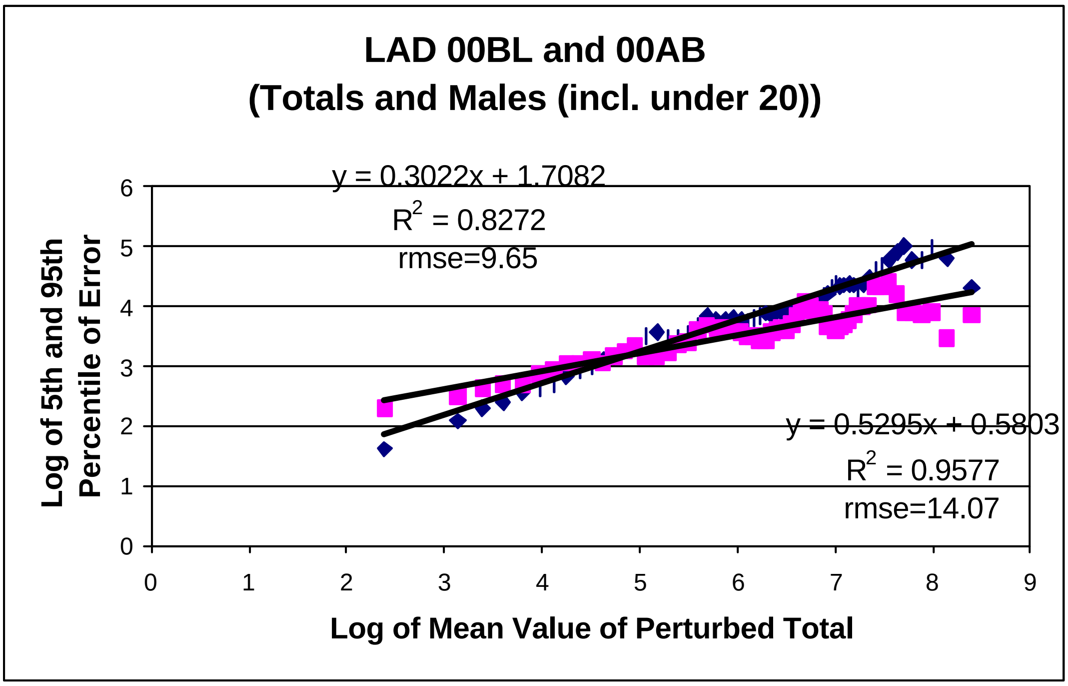



Based on the results as seen in the tables of the Appendix and the above figures, for the linear regression model using the square root of the mean value of the perturbed totals as the independent variable, we see inconsistent estimates for the regression coefficients between the two LAD's. For example, for the 5$^{th}$ percentile, the intercepts of the models for MG201_00BL and MG201_00AB are –7.459 and 16.952, respectively. Therefore, it would be more difficult pooling different LAD's into one consistent model for predicting the bounds of the perturbation interval. The regression model that has overall the smallest root mean square error of the predicted values of the percentiles of the error distribution and the largest $R^2$ is the multiplicative model. The regression coefficients are sufficiently consistent so that different LAD's and different sub-populations may be combined to obtain one set of estimates for the intercept and slope for each of the bounds of the perturbation interval.

A separate analysis was carried out for perturbed totals under 20. In spite of obtaining different slopes and intercepts and an overall better fit for the perturbation interval for this group, the difference between the predicted bounds when including them in the general regression analysis was not large and tended to provide a larger width for the perturbation interval for this group. Therefore, in order to simplify the guidelines for users, the final regression analysis will include an additional band $k$ for the perturbed totals up to 20.

As an example of pooling the two LAD totals and sub-totals of the migration table M201 as well as including the perturbed totals less than 20, Figure 4 presents the multiplicative model for all of the tables.

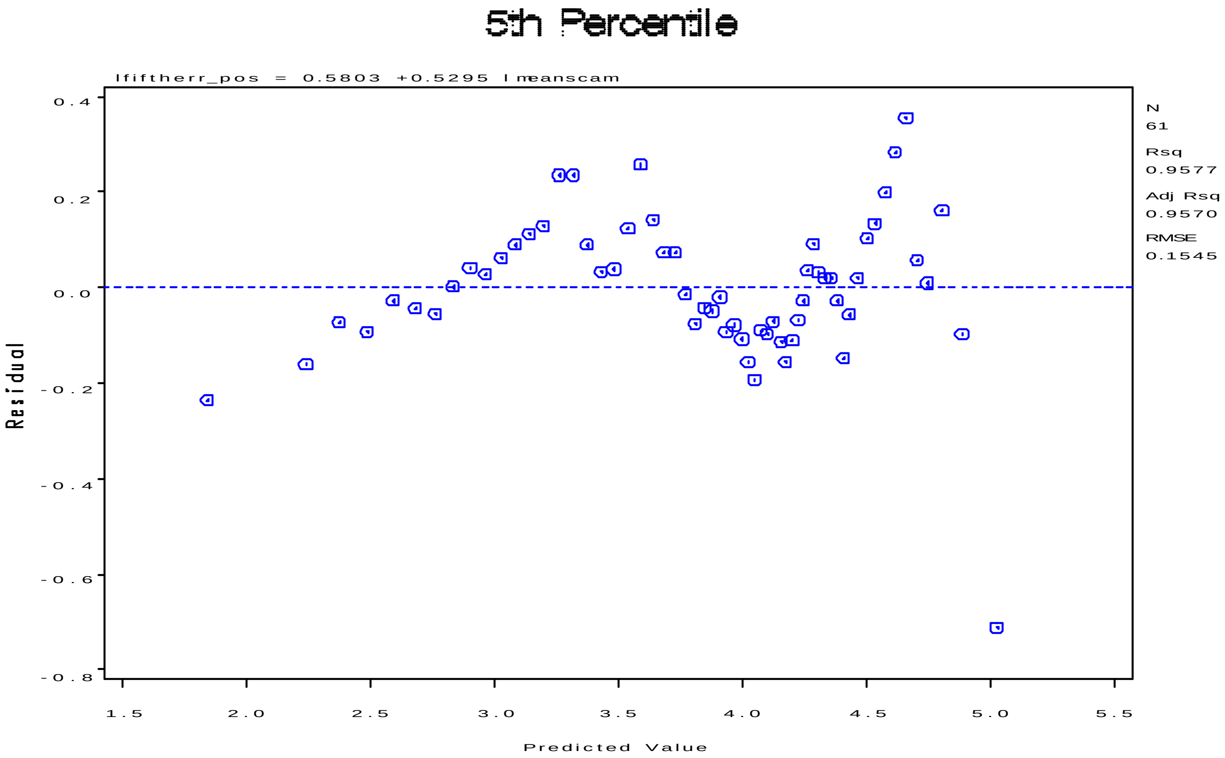
**Figure 4: Regression of Log of Percentiles of Errors on Log of Mean Value of Perturbed Totals– Migration Tables MG201_00BL and MG201_00AB (Totals, Males)**

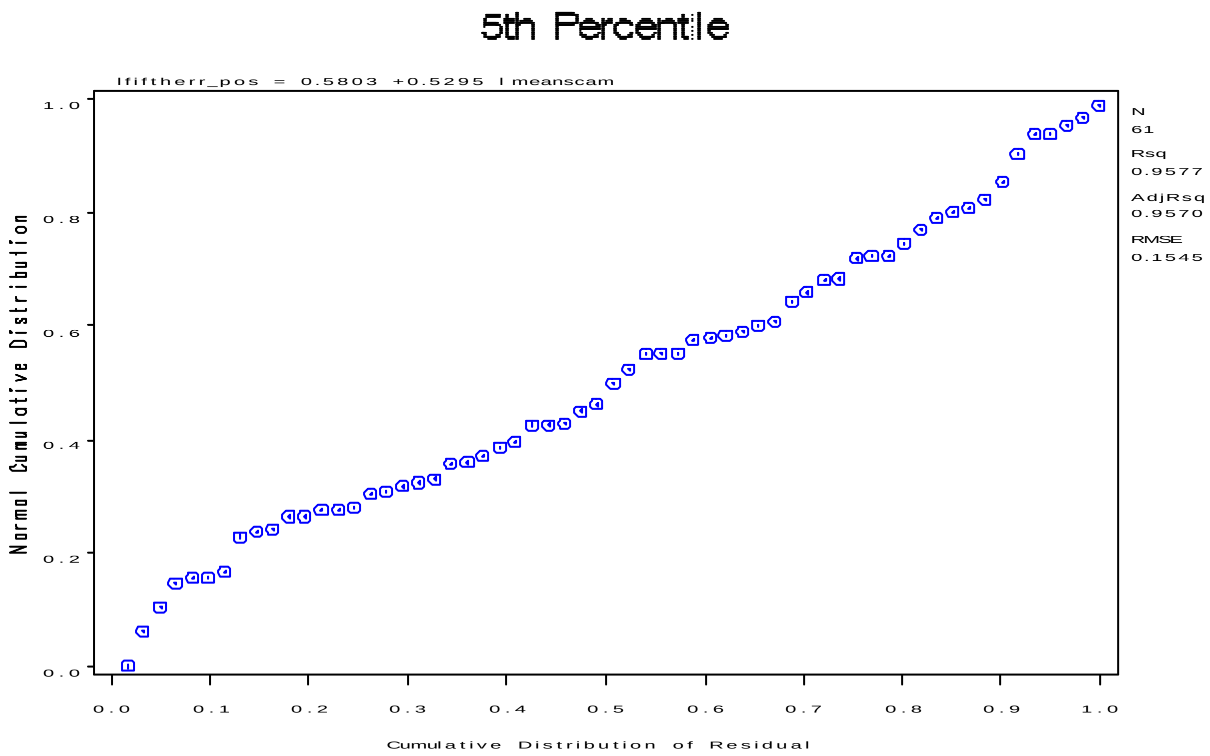



Recall that the groupings that were generated from the LAD tables are dependent since an origin ward is included in many of the groupings of various sizes and therefore the residuals are correlated. To show this we plot in Figure 5 the predicted values against the residuals as well as the normal probability plot for testing whether the residuals are normally distributed with a mean of zero. These plots are based on the regression model in Figure 4 where the dependent variable is the log of the $5^{th}$

percentile and the independent variable is the log of the mean value of the perturbed total.

**Figure 5: Analysis of Residuals for Regression of Log of Percentiles of Errors on Log of Mean Value of Perturbed Totals for Migration Tables**

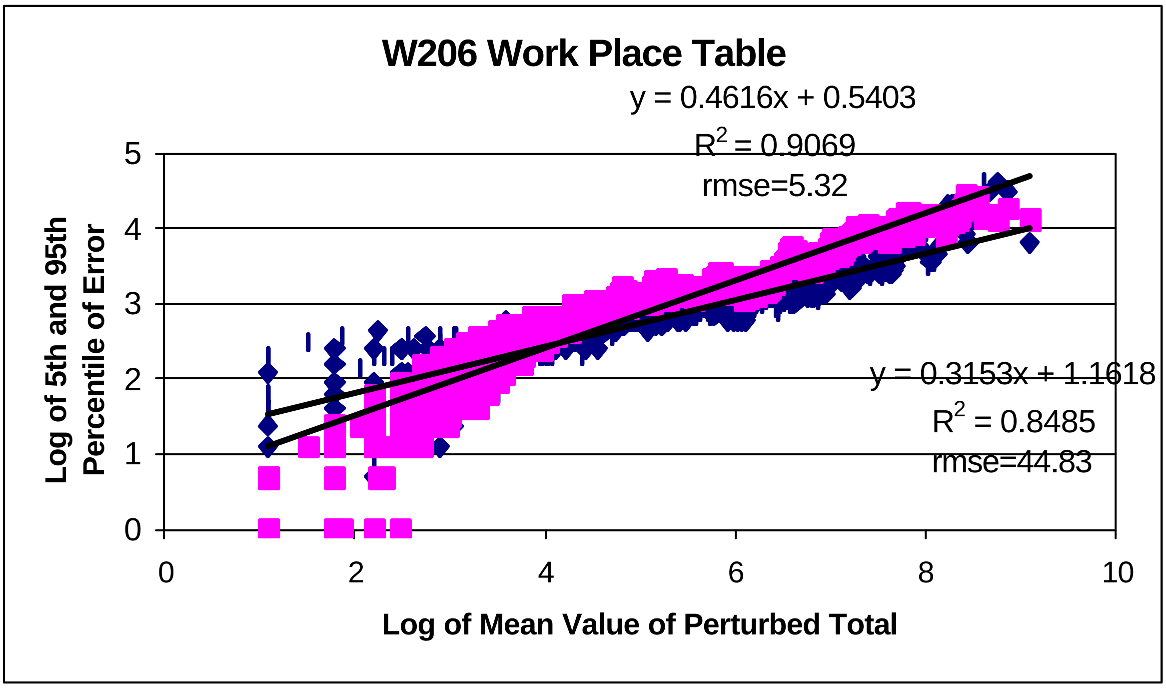


The Kolmogorov-Smirnov statistic for normality is 0.1151 with a p-value of 0.0437 thus the hypothesis is rejected. Since the residuals are correlated this result is expected. Note that the purpose here is not to make statistical inferences rather to provide a best linear fit for a set of points so that users may calculate a perturbation interval that will cover approximately 90% of the errors. This method of ordinary least squares used in regression modelling which calculates an intercept and a slope provides the best linear fit.

In Figure 6 we present the multiplicative regression model for the workplace table W206, across all the LAD's in West Midlands for totals, males and females.

**Figure 6: Regression of Log of Percentiles of Errors on Log of Mean Value of Perturbed Totals – Workplace Table W206 for West Midlands (Totals, Males, Females)**



## 5. Recommendations

To provide users with guidelines for assessing the extent of the errors caused by the perturbation scheme when aggregating up from perturbed lower geographical origin-destination tables, we recommend providing formulae based on the multiplicative model for each of the predicted bounds of the perturbation interval. This interval will cover approximately 90% of the expected differences from the perturbed raw totals. The predicted lower bound for the 90% perturbation interval based on the regression model for the $5^{th}$ percentile, (i.e., the intercept $a_L$ and the slope $b_L$) and the aggregated perturbed total $T^p$ is: $T^p - a'_L (T^p)^{b_L}$ where $a'_L = \exp(a_L)$. The predicted upper bound for the 90% perturbation interval based on the regression model for the $95^{th}$ percentile is: $T^p + a'_U (T^p)^{b_U}$ . The exact values of the intercepts and slopes for the formulae will be based on averaging or pooling the different regression models as shown in Figures 4 and 6. For example, based on the regression model of Figure 4 for the migration table, the formulae are as follows:

- Upper bound of prediction interval: $a_U = 1.708$, $\exp(a_U) = 5.519, b_U = 0.302$,
  $T^p + \exp(a_U)(T^p)^{b_U} = T^p + 5.519(T^p)^{0.302}$
- Lower bound of prediction interval: $a_L = 0.580$, $\exp(a_L) = 1.787, b_L = 0.530$,
  $T^p - \exp(a_L)(T^p)^{b_L} = T^p - 1.787(T^p)^{0.530}$

and based on the regression model of Figure 6 for the workplace table, the formulae are as follows:

- Upper bound of prediction interval: $a_U = 0.540$, $\exp(a_U) = 1.717, b_U = 0.462$,
  $T^p + \exp(a_U)(T^p)^{b_U} = T^p + 1.717(T^p)^{0.462}$
- Lower bound of prediction interval: $a_L = 1.162$, $\exp(a_L) = 3.196, b_L = 0.315$,
  $T^p - \exp(a_L)(T^p)^{b_L} = T^p - 3.196(T^p)^{0.315}$

It is highly recommended that other tables be examined to expand the analysis and obtain a better fit for the perturbation interval. This analysis is based solely on a few specific tables and therefore it is not possible at this stage to derive one generalised perturbation interval for every migration or workplace table. Efforts will be made in the future to expand the analysis to other tables and sub-groups of the population.

The formulae should be applied to the contribution of ward data to the total. For example, when totals contain both ward level data and LAD level data, a perturbation interval should be calculated for the ward level data as explained above and then the LAD level data should be added on to each of the bounds of the perturbation interval.

In addition, to assist users that would want to benchmark or prorate their data for statistical analysis, it would be useful to provide simple tables comparing the previously published higher geographical level totals and sub-population totals with the aggregated perturbed totals from the lower geographical origin-destination tables. Also, origin-destination tables at all geographical levels should be produced for the total population or main sub-groups of the population for each table without further disaggregation to allow for minimum differences from the raw totals.

Besides giving formulae for calculating the predicted bounds of the perturbation interval, a graph should be provided which would allow users to immediately assess the extent of the differences from the raw totals for any perturbed total without having to carry out calculations. The x-axis contains the perturbed totals. The predictions of the $5^{th}$ and $95^{th}$ percentile error bounds from the regression models are plotted and smoothed on the y-axis. Based on the graph, the user can assess for every perturbed ward total, the approximate 90% perturbation interval. Figures 7 and 8 present examples of a graph based on the multiplicative regression model for the combined migration tables MG201_00BL and MG201_00AB and the workplace table W206 for West Midlands as shown in Figures 4 and 6. In addition, small tables may be useful to the users that would provide bounds of the prediction intervals for specified rounded perturbed totals as shown in Tables 3 and 4.

**Figure 7: Perturbation Interval Bounds According to Perturbed Total for Origin-Destination Migration Table MG201**

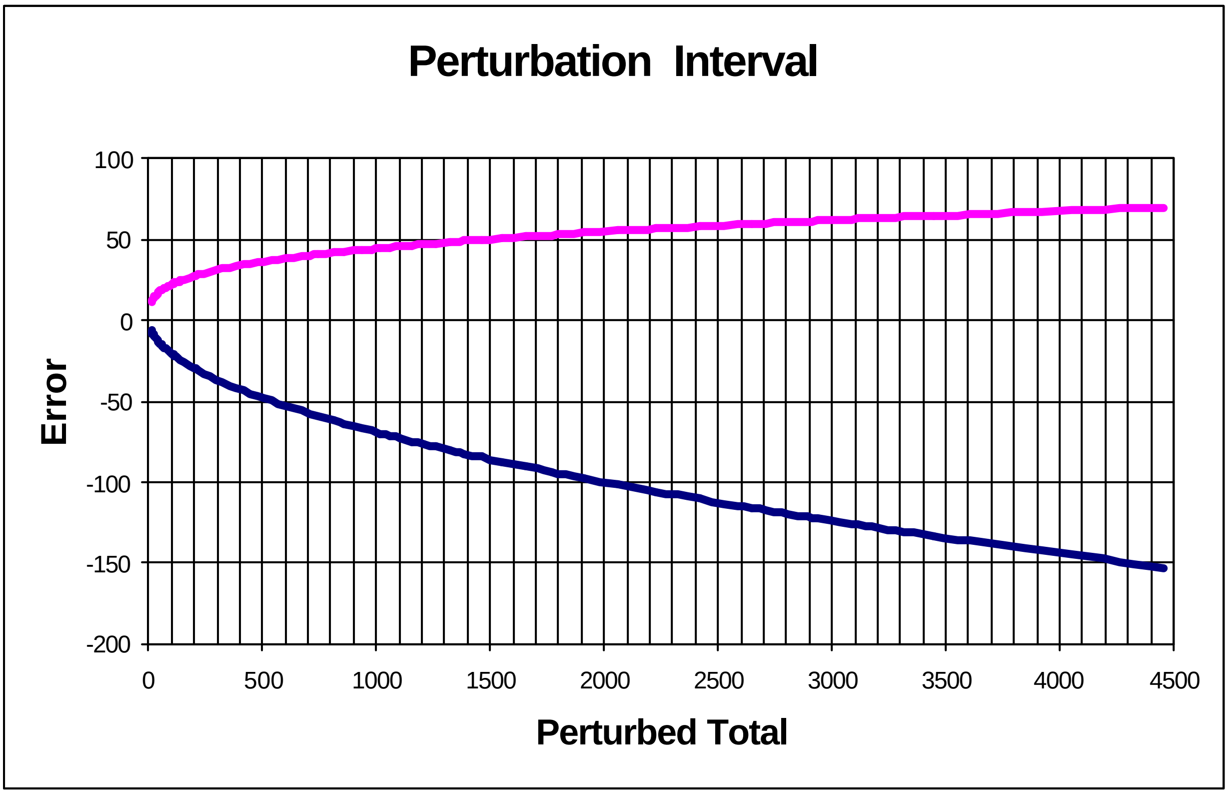


**Figure 8: Perturbation Interval Bounds According to Perturbed Total for Origin-Destination Workplace Table W206**

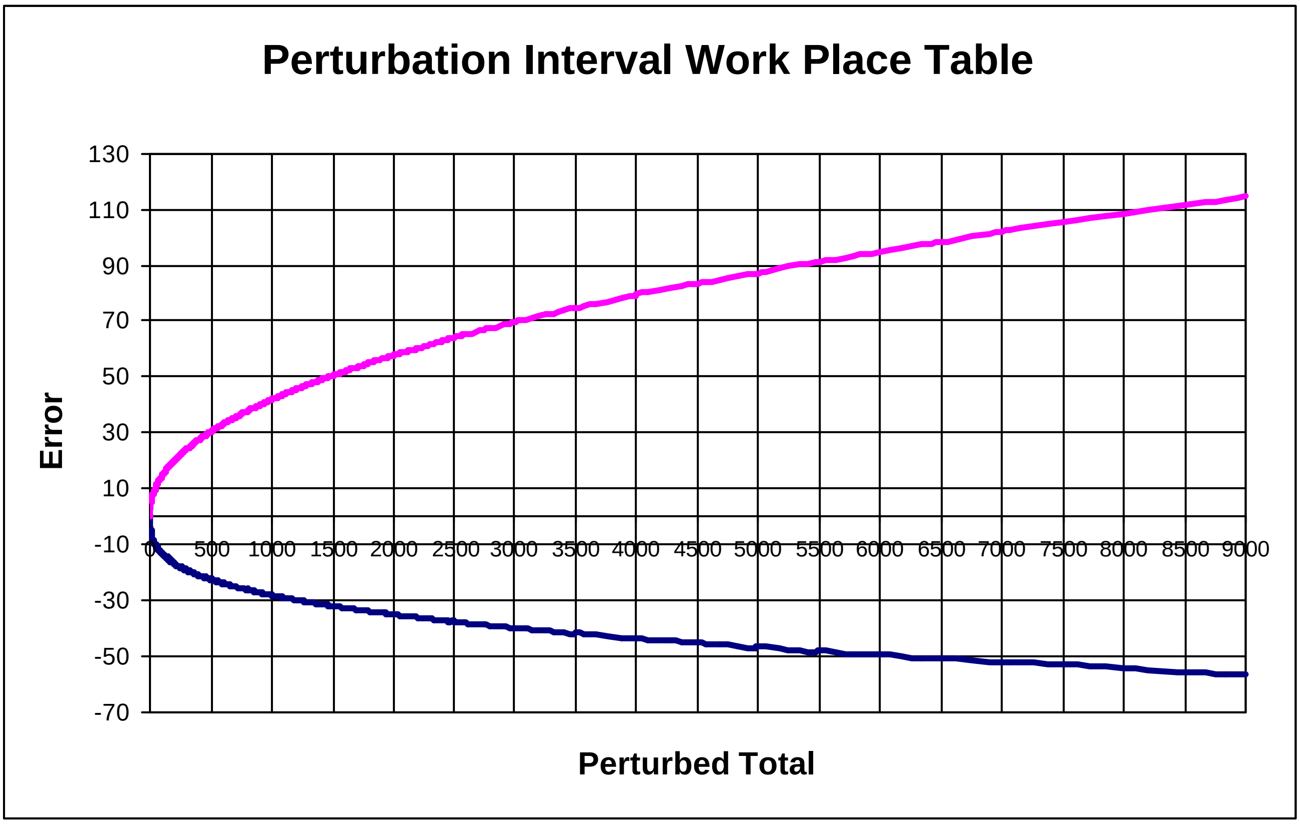


**Table 3: Prediction Interval for Specified Perturbed Totals for Migration Table MG201**

| Perturbed Total | Prediction Interval | | Perturbed Total | Prediction Interval | |
|---|---|---|---|---|---|
| | Lower Bound | Upper Bound | | Lower Bound | Upper Bound |
| 5 | 1 | 14 | 380 | 339 | 413 |
| 10 | 4 | 21 | 400 | 357 | 434 |
| 20 | 11 | 34 | 420 | 376 | 454 |
| 30 | 19 | 45 | 440 | 395 | 475 |
| 40 | 27 | 57 | 460 | 414 | 495 |
| 50 | 36 | 68 | 480 | 433 | 516 |
| 60 | 44 | 79 | 500 | 452 | 536 |
| 70 | 53 | 90 | 550 | 500 | 587 |
| 80 | 62 | 101 | 600 | 547 | 638 |
| 90 | 71 | 111 | 650 | 595 | 689 |
| 100 | 80 | 122 | 700 | 643 | 740 |
| 120 | 97 | 143 | 750 | 691 | 791 |
| 140 | 116 | 165 | 800 | 738 | 842 |
| 160 | 134 | 186 | 850 | 786 | 892 |
| 180 | 152 | 207 | 900 | 834 | 943 |
| 200 | 170 | 227 | 950 | 883 | 994 |
| 220 | 189 | 248 | 1,000 | 931 | 1,045 |
| 240 | 207 | 269 | 1,500 | 1,414 | 1,550 |
| 260 | 226 | 290 | 2,000 | 1,900 | 2,055 |
| 280 | 245 | 310 | 2,500 | 2,387 | 2,559 |
| 300 | 263 | 331 | 3,000 | 2,876 | 3,062 |
| 320 | 282 | 352 | 3,500 | 3,366 | 3,565 |
| 340 | 301 | 372 | 4,000 | 3,856 | 4,068 |
| 360 | 320 | 393 | 4,500 | 4,346 | 4,570 |

**Table 4: Prediction Interval for Specified Perturbed Totals for Workplace Table W206**

| Perturbed Total | Prediction Interval | | Perturbed Total | Prediction Interval | |
|---|---|---|---|---|---|
| | Lower Bound | Upper Bound | | Lower Bound | Upper Bound |
| 5 | 0 | 9 | 500 | 477 | 530 |
| 10 | 3 | 15 | 550 | 527 | 582 |
| 20 | 12 | 27 | 600 | 576 | 633 |
| 30 | 21 | 38 | 650 | 625 | 684 |
| 40 | 30 | 49 | 700 | 675 | 735 |
| 50 | 39 | 60 | 750 | 724 | 786 |
| 60 | 48 | 71 | 800 | 774 | 838 |
| 70 | 58 | 82 | 850 | 823 | 889 |
| 80 | 67 | 93 | 900 | 873 | 940 |
| 90 | 77 | 104 | 950 | 922 | 991 |
| 100 | 86 | 114 | 1,000 | 972 | 1,042 |
| 120 | 106 | 136 | 1,250 | 1,220 | 1,296 |
| 140 | 125 | 157 | 1,500 | 1,468 | 1,550 |
| 160 | 144 | 178 | 1,750 | 1,716 | 1,804 |
| 180 | 164 | 199 | 2,000 | 1,965 | 2,057 |
| 200 | 183 | 220 | 2,250 | 2,214 | 2,311 |
| 220 | 202 | 241 | 2,500 | 2,462 | 2,564 |
| 240 | 222 | 262 | 2,750 | 2,711 | 2,816 |
| 260 | 242 | 282 | 3,000 | 2,960 | 3,069 |
| 280 | 261 | 303 | 3,500 | 3,458 | 3,574 |
| 300 | 281 | 324 | 4,000 | 3,956 | 4,079 |
| 320 | 300 | 345 | 4,500 | 4,455 | 4,583 |
| 340 | 320 | 365 | 5,000 | 4,953 | 5,087 |
| 360 | 340 | 386 | 5,500 | 5,452 | 5,591 |
| 380 | 359 | 407 | 6,000 | 5,950 | 6,095 |
| 400 | 379 | 427 | 6,500 | 6,449 | 6,599 |
| 420 | 399 | 448 | 7,000 | 6,948 | 7,102 |
| 440 | 418 | 468 | 7,500 | 7,447 | 7,606 |
| 460 | 438 | 489 | 8,000 | 7,946 | 8,109 |
| 480 | 458 | 510 | 9,000 | 8,944 | 9,115 |

## Appendix

**Table 1: Summary of Regression Models on LAD MG201_00BL for 5th Percentile**

| 5th Percentile | Independent Variable | Intercept | Slope | $R^2$ | Root MSE |
|---|---|---|---|---|---|
| **Totals:** | | | | | |
| Error | Perturbed | -25.402 | -0.017 | 0.638 | 16.80 |
| Error | Sqrt. Perturbed | -7.459 | -1.281 | 0.711 | 15.00 |
| Log Error | Log Perturbed | 1.125 | 0.397 | 0.846 | 15.08 |
| **Males:** | | | | | |
| Error | Perturbed | -20.098 | -0.018 | 0.522 | 10.56 |
| Error | Sqrt. Perturbed | -9.617 | -0.993 | 0.623 | 9.39 |
| Log Error | Log perturbed | 1.242 | 0.356 | 0.697 | 9.27 |

**Table 2: Summary of Regression Models on LAD MG201_00AB for 5th Percentile**

| 5th Percentile | Independent Variable | Intercept | Slope | $R^2$ | Root MSE |
|---|---|---|---|---|---|
| **Totals:** | | | | | |
| Error | Perturbed | -21.572 | -0.060 | 0.959 | 8.53 |
| Error | Sqrt. Perturbed | 16.952 | -3.343 | 0.907 | 12.90 |
| Log Error | Log Perturbed | 0.509 | 0.570 | 0.942 | 12.84 |
| **Males:** | | | | | |
| Error | Perturbed | -17.412 | -0.043 | 0.777 | 7.95 |
| Error | Sqrt. Perturbed | -2.154 | -1.793 | 0.828 | 6.98 |
| Log Error | Log perturbed | 0.797 | 0.472 | 0.848 | 7.01 |

**Table 3: Summary of Regression Models on LAD MG201_00BL for 95th Percentile**

| 95th Percentile | Independent Variable | Intercept | Slope | $R^2$ | Root MSE |
|---|---|---|---|---|---|
| **Totals:** | | | | | |
| Error | Perturbed | 27.579 | 0.006 | 0.369 | 9.67 |
| Error | Sqrt. Perturbed | 19.463 | 0.490 | 0.548 | 8.19 |
| Log Error | Log Perturbed | 1.926 | 0.240 | 0.777 | 7.44 |
| **Males:** | | | | | |
| Error | Perturbed | 17.708 | 0.018 | 0.820 | 5.19 |
| Error | Sqrt. Perturbed | 8.587 | 0.933 | 0.860 | 4.58 |
| Log Error | Log perturbed | 1.553 | 0.298 | 0.874 | 5.04 |

**Table 4: Summary of Regression Models on LAD MG201_00AB for 95th Percentile**

| 95th Percentile | Independent Variable | Intercept | Slope | $R^2$ | Root MSE |
|---|---|---|---|---|---|
| **Totals:** | | | | | |
| Error | Perturbed | 32.916 | 0.025 | 0.391 | 21.43 |
| Error | Sqrt. Perturbed | 11.253 | 1.592 | 0.488 | 19.66 |
| Log Error | Log Perturbed | 1.156 | 0.425 | 0.668 | 19.66 |
| **Males:** | | | | | |
| Error | Perturbed | 25.419 | 0.020 | 0.462 | 7.40 |
| Error | Sqrt. Perturbed | 16.173 | 0.950 | 0.648 | 5.99 |
| Log Error | Log perturbed | 1.756 | 0.306 | 0.786 | 5.60 |