Benchmark minor road traffic flow estimates for England and Wales, 2008-09

Charles Lound, *Office for National Statistics,* October 2011

Introduction

The Department for Transport have estimated minor road traffic flow estimates for Great Britain using a sample of count points which remained fixed over a period of around a decade. Retaining a fixed sample provides more precise measures of change, so we will refer to this as the *change sample*. However, the error associated with accumulated change estimates can lead to *rolled forward* traffic estimates with substantial random error and, potentially, systematic error.

An independent, large benchmark sample of count points was drawn and counted during 2008 and 2009 in parallel with the annual enumeration of the change sample. This benchmark sample was stratified by local authority and a version of road category, allocated to strata using recent traffic estimates, and drawn with probability proportional to the length of the minor road segment. This note shows how the benchmark count data can be used to produce benchmark estimates of minor road traffic for 2008-09 which can be used for comparison with the existing estimates and to make adjustments to provide a consistent series. The estimate is derived by calibrating a standard design-weighted estimator to published road length totals.

The sample design

The documentation and discussion of the benchmark sample has confirmed that the original sample was close to the intended random sample of minor road TOIDs[1], stratified by local authority and road category and drawn with probability proportional to the TOID length. The departures from the intended random sampling in the original sample were as follows.

- The most obvious omission is that none of the Scottish benchmark points were actually counted. This in effect limits the benchmark estimation to England and Wales, but leaves hanging the issue of how best to estimate for Scotland at the benchmark point.
- The original frame included only those TOIDs that were described as single or dual carriage roads, omitting roundabouts, slip roads etc. This means that the frame included 99% of the minor road network by length, so omitting these other sections will have a minimal impact which can be mitigated by weighting.

[1] TOID is short for Topographic Identifier allocated by the Ordnance Survey to every feature on their MasterMap® product. Here we use it to refer to those TOIDS that identify minor roads.

- The sample was reportedly allocated using estimated stratum traffic. After rounding, 178 of the 1,152 strata had a benchmark sample allocation of zero and so had zero chance of contributing to the estimate. The omitted strata cover only 2% of the road network by length on the frame. Some compensation for this can be included in a weighting adjustment during estimation.

In addition, the following changes were made to the original sample which then further departs from random sampling:

- Benchmark sample points that had already been sampled for the existing change sample were substituted for points which were judged to be similar. It is unclear how many such points were substituted, but as the sampling rates for the two samples are 1-in-252 for the benchmark sample and 1-in-634 for the change sample then we expect the number will be minimal. The biasing effect of the substitution will depend on the quality of the judgement in replacing with a similar point but it seems reasonable to assume this has worked well.
- Benchmark sample points that were close to another sampled point were also moved in a similar way. This was also applied to sections of road that would have been difficult to count. It would be difficult or impossible to audit this in detail but again it seems reasonable to assume this has worked well.
- Low flow sites were identified in advance of counting. In 2008, two measures were taken for these. The count point could have been moved to the entrance of a cul-de-sac. Alternatively, to help estimate the impact of such a modification, the original count point could be retained together with a second count nearby. It is also important to note that the road categories used to stratify the benchmark sample, although labelled with the standard cross-classification of B/C/unclassified road type by urban/rural are *not* comparable with this standard classification. See later section on *Calibration level* for more discussion on this.

Estimation method

Each sampled TOID has an allocated count point near its centre. Any such count point is one of the theoretically infinite number of points that could be taken along the length of the road, each of which could differ because of traffic entering or exiting part way along the road. It seems reasonable to assume that on average across all roads the centre-point count is close to the average count taken across all possible count points.

For each sampled TOID *k* in stratum *h* we have a count y_{hk} for a vehicle type (or all vehicles) and a TOID length l_{hk}. There were n_h such TOIDs sampled from the stratum.

From a separate exercise using Automatic Traffic Counters, we have an expansion factor f_{ed}, for the expansion group *e* containing the current TOID *(hk)* and the day, *d*, that the count was taken. (See Appendix A for a brief description of the derivation of the expansion factors.) This is used to translate the twelve-hour daily count into an estimated Annual Average Daily Flow (AADF) for each TOID.

$$a_{hk} = f_{ed} y_{hk}$$

If we multiply this daily average flow by the length of the TOID and the number of days in the year we can estimate the total traffic for that TOID in a year. We can then use this in a standard design-weighted estimator to estimate the total traffic for the population:

$$\hat{t} = \sum_{hk} d_{hk} (Y_{hk} a_{hk} l_{hk})$$

where d_{hk} is a design weight for each sampled case and Y_{hk} is the number of days in the year the count was taken (either 365 or 366). Since the original sampling within strata was with probability proportional to TOID length, the design weights are

$$d_{hk} = \frac{\Lambda_h}{n_h \lambda_{hk}}, \quad n_h > 0$$

where λ_{hk} was the TOID length at the time of sampling and Λ_h the population stratum TOID length at the time of sampling. The fact that this can be created only for strata with some sample emphasises the problem noted earlier about strata which had zero samples. The design weight also accounts for the over-sampling by previous estimates of traffic, through the sample allocation n_h, as strata with larger sample allocations will get a lower weight and vice versa.

Note that the time lag between sampling and counting for the benchmark counts is short so the sampled road lengths will typically be unchanged. If so, and assuming a constant year length Y then the above estimator becomes:

$$\hat{t}' = \sum_{hk} \frac{\Lambda_h}{n_h \lambda_{hk}} (Y a_{hk} \lambda_{hk}) = \sum_h Y \Lambda_h \bar{a}_h \text{ where } \bar{a}_h = \sum_k a_{hk} \Big/ n_h$$

Therefore under this simplification, the proposed standard estimator is the same as the one previously proposed during initial work on the benchmark sample. There the traffic estimate for each stratum was created by taking the unweighted mean of the AADFs and multiplying that by the total stratum road length and the year length.

The usual next step in the estimation process is to seek to improve the design weighted estimator by calibrating the weights to auxiliary data available for each sampled element and the whole population. Road length is the obvious auxiliary variable, for which the calibration condition is:

$$\sum_{(hk) \in c} w_{hk} l_{hk} = L_c, \; c = 1 \ldots C$$

for a given set of classes. This can be extended to more than one set of classes, simultaneously covering more than one dimension.

In fact, though, if the road lengths haven't changed then, from the formula for the design weights above, $\sum_{k \in h} d_{hk} l_{hk} = L_h$, the full set sample is self-calibrating at the stratum level and this will apply to any classes comprising whole strata. However, this condition is true for the whole set sample and in practice we do not have counts for the whole set sample because of the departures from the random sampling described previously and other dropped count points described in the following section. Therefore we can use this calibration stage to modify the design weights to attempt to reduce any biasing effect from some of the set sample not being counted.

Linking the sample and count data

The standard estimator requires each count to be associated with the original sampling unit and therefore the design weight. The original sample was taken into the standard system for managing counts and so a set of AADFs has been produced for a benchmark sample with a set of count point codes.

In principle, establishing the required link needed a lookup from the sample TOID code to the CP count point code in use. Since some of the CP codes had been re-coded to new values, this in practice required more than one step and careful cross-referencing within the matched sample to check this has worked. (The details, including the tables involved are shown in appendix B.) Where a count point could be allocated to the sampled unit, a further data table could be used to establish that some of the sampled count points had been closed. Following this matching the original sample could be accounted for as shown in Table 1.

Table 1: Accounting for the original benchmark sample

		Sample units	%
AADF count linked		9,729	*93*
Count Point linked but closed	616	754	*7*
No Count Point linked	138		
Total England and Wales		10,483	*100*

Of the 616 count points that were linked but closed, 569 could be accounted for in administrative records as follows:

a) 172 were cul de sacs;

b) 4 had health and safety issues preventing a manual count taking place;

c) 96 were in Wales and were closed due to the limited funding available for the counts;

d) 2 should have been rescheduled rather than closed;

e) 72 matched the location of another count point;

f) 118 were on a private road or a dirt track etc; and

g) 5 had been re-classified due to network changes;

The first five groups can be seen as units which should have been counted. However, it appears that the 96 count points withdrawn for Wales were withdrawn as a random sub-sample of the whole sample so, while their absence will increase the variance of traffic estimates, it will not cause bias. Group e) may merit further investigation because where the count point has been counted as part of the change sample, those counts can legitimately be used. The last two groups are not of concern as they are ineligible for the minor road traffic estimates.

Calibration level

To calibrate to road length estimates we need comparable classifications that can be applied to the benchmark sample count points and road length data. We would also like to identify classifications that explain the variation in the rate at which count points were omitted and also explain variation in traffic levels.

Minor roads can be broken down by type into B, C or unclassified. These are reflected in the labelling of the sample strata, but in fact the C/unclassified distinction was not directly coded in the source data and was applied using a proxy from the road descriptors[2]. Therefore, we can expect the B *versus* breakdown to be comparable with other sources, but not the complete three-way split.

Similarly a GIS-based comparison of the urban/rural split applied to stratify the sample does not match with the standard definition which is based on location within an urban area with a population of size 10,000 or more.

In view of these reservations we have chosen to carry out the calibration to road lengths by breaking down the sample by region (directly coded from the local authority) by road type classified into B *versus* [C/unclassified]. Since this is carried out in a complete cross-classification, the process consists simply of *grossing* each traffic estimate within each region by road type cell by the ratio of the road length from published estimates to the estimated road length within the sample.

The cell sizes for the calibration are shown in appendix C. These are clearly all larger than the usual rule-of-thumb that cells sizes should be no smaller than thirty cases, and suggest that breaking into a finer classification would be possible, although would require identification of consistent subgroup variables with associated totals. (Breaking down as far as local authority would lead to too small cells.)

Benchmark traffic estimates

Using the linked data, we have applied the standard design-weighted estimator to create sample estimates for traffic for all vehicles and road lengths. These are shown in Table 2 disaggregated by region and separately for B roads and for other minor roads.

The table also shows published road lengths, averaged for 2008 and 2009. These are somewhat larger than the design-weighted estimates of road length. This reflects the small amount of undercoverage in the frame, noted earlier, but more importantly the fact that not all the sampled count points were in fact enumerated. The table shows how the ratio of the published road lengths to their design-weighted counterparts has been used to weight up the traffic estimates. The calibration factors shown on the right of the table are simply the road length factors used in this weighting up, showing the extent to which each design-weighted traffic estimate was increased by the calibration.

[2] The breakdown of road length into C and unclassified roads is achieved through the annual consultation with local authorities. These returns give an aggregate breakdown into the classes but are not used to code individual roads on the database.

The most obvious distinction in the calibration factors is between England where the factors range from 1.02 to 1.13 and Wales with factors of 1.64 and 1.72. This is to be expected as it relates to, and properly accounts for, the random subsampling of count points in Wales.

Table 2: Calibrating the benchmark estimates of traffic using known road lengths by region and broad road type

	Design-weighted estimates					Published estimates		Calibrated estimator			Calibration factors	
		Traffic (bnvkm)		Road lengths (km)		Road lengths (km)		Traffic (bnvkm)				
	All roads	B roads	Other minor roads	B roads	Other minor roads	B roads	Other minor roads	B roads	Other minor roads	All roads	B roads	Other minor roads
		[a]	[b]	[c]	[d]	[e]	[f]	[a].[e]/[c]	[b].[f]/[d]		[e]/[c]	[f]/[d]
North East	7.0	1.7	5.4	1,264	12,810	1,338	13,040	1.8	5.4	7.2	1.06	1.02
North West	17.1	4.0	13.0	1,797	29,805	1,981	30,320	4.4	13.3	17.7	1.10	1.02
Yorks. & the Humber	14.2	3.4	10.9	1,814	25,434	1,916	26,444	3.6	11.3	14.9	1.06	1.04
East Midlands	13.5	3.4	10.1	1,912	23,938	2,011	25,190	3.6	10.6	14.2	1.05	1.05
West Midlands	17.2	4.9	12.3	2,440	25,654	2,507	26,356	5.1	12.6	17.7	1.03	1.03
East of England	19.9	5.6	14.3	3,032	30,402	3,246	32,219	6.0	15.2	21.2	1.07	1.06
London	10.2	1.8	8.4	449	12,071	509	12,509	2.0	8.7	10.7	1.13	1.04
South East	27.0	7.0	20.0	2,928	37,602	3,075	38,435	7.3	20.5	27.8	1.05	1.02
South West	18.4	5.0	13.4	3,056	40,105	3,265	41,420	5.4	13.8	19.2	1.07	1.03
Wales	5.9	1.8	4.1	1,823	15,465	2,989	26,560	2.9	7.0	10.0	1.64	1.72
England	144.6	36.8	107.8	18,693	237,821	19,848	245,931	39.2	111.4	150.6		
England and Wales	150.5	38.6	111.9	20,515	253,286	22,838	272,491	42.1	118.5	160.6		

Providing a consistent traffic series for England and Wales

The above results provide a set of benchmark estimates for comparison with the rolled-forward estimates. This comparison, at a regional level, is shown in Table 3, with the column for the rolled forward estimates showing the mean of the 2008 and 2009 published figures.

The final column shows the proposed adjustment to the rolled forward estimates, calculated as the ratio of the two sets of estimates. This could be applied as a single factor for England and Wales as a whole, or as separate factors for each region.

Table 3: Estimated traffic by region, rolled forward and benchmark estimates

Region	Rolled forward traffic estimate, mean 2008 & 09 (bnvkm)	Benchmark traffic estimate, 2008-09 (bnvkm)	*Benchmark sample sizes (achieved)*	Potential adjustment to benchmark
	[A]	[B]		[B/A]
North East	9.0	7.2	*526*	0.81
North West	19.4	17.7	*1,118*	0.91
Yorkshire and the Humber	15.2	14.9	*891*	0.98
East Midlands	13.8	14.2	*792*	1.03
West Midlands	19.0	17.7	*1,150*	0.93
East of England	21.7	21.2	*1,290*	0.98
London	12.6	10.7	*740*	0.85
South East	29.9	27.8	*1,805*	0.93
South West	17.8	19.2	*1,029*	1.08
Wales	10.5	10.0	*386*	0.95
England and Wales	168.8	160.6	*9,727*	0.95

Source: Table TRA0203

A more formal description of the estimators involved and their variances is shown in appendix D. However, we do not have the variance estimates that would be required to make a decision based on overall variance.

The overall variance is in part dependent on the variance of the benchmark estimates. The large overall sample size would appear to suggest the overall benchmark estimate is reliable. However, applying an overall adjustment loses some of the detail shown by the variation in the regional adjustments, from 0.81 in the North East to 1.09 in the South West which may include some systematic regional effects. Since sample sizes remain substantial at the regional level, we judge that it would be appropriate to disaggregate the calculation down to this level. Going further, down to, say, local authority level would result in small sample sizes and volatile benchmark estimates which would negate some of the benefit of benchmarking.

In addition to breaking the adjustments down geographically, we also explored applying different benchmarking factors by vehicle type as shown in Table 4. The comparison for all motor vehicles is as shown in the regional breakdown with a potential reduction of 5% on adjustment to the benchmark. This is largely dependent on the values for cars and taxis which dominate these traffic totals. For HGVs the potential adjustment would be to increase the traffic estimates by 13%, but for all the other motor vehicles there would be substantial reductions of between 25% and 35%.

Table 4: Estimated traffic by vehicle type, rolled forward and benchmark estimates

Vehicle type	Existing traffic estimate mean 2008 & 2009	Benchmark traffic estimate, 2008-09 (bnvkm)	Potential adjustment to benchmark
	[C]	[D]	[D]/[C]
Two-wheeled motor vehicles	2.5	1.6	0.65
Cars & taxis	136.6	135.3	0.99
Buses & coaches	2.4	1.9	0.81
Light goods vehicles	24.3	18.3	0.75
All HGVs	3.0	3.4	1.13
All Motor vehicles	168.8	160.6	0.95
Pedal cycles	*3.8*	*2.4*	*0.64*

We might expect that some of these larger potential adjustments are due to the fact that these vehicle types are relatively rare within the vehicle counts which leads to a high relative variance but, without a formal indication of the variance, it is difficult to make an assessment of this volatility. We do not have any external evidence qualitative or quantitative to suggest that such adjustments would lead to a likely reduction in bias.

In looking for further justification for the form of an adjustment, we can look at the form of the estimator used to produce the rolling estimates. In those calculations, year-on-year growth factors for traffic are calculated. These growth factors are calculated separately for rural and urban areas and for each vehicle type, but not separated by region. Therefore from this process we might expect the estimates to reflect different growths in traffic by vehicle type but not by region. This appears to add to the justification for adjusting by region, but not by vehicle type.

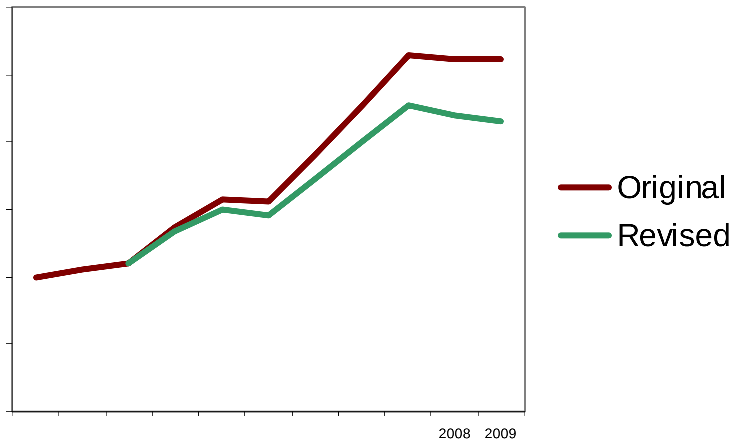
Therefore our recommendation is to make the benchmarks adjustments at the regional level.

Providing a consistent back series

The adjustment to the benchmark estimates is intended to correct for accumulated bias and variance from when the change sample was first introduced. We therefore propose to adjust the back series using annual regional factors that converge to one at the start of the change sample. Since the adjustments are applied as a factor, the annual adjustment factors are in a geometric series, so that if the regional adjustment factor at the benchmark point is γ_r, then the adjustment factor for t years before the benchmark is $\gamma_r^{\frac{T-t}{T}}$, where T is the number of years between the start of the change sample and the benchmark point.

The impact of this is illustrated in the following diagram representing an artificial original and adjusted set of estimates:

Figure 1: Illustration of the impact of benchmark adjustment with tapering



We recommend that the same region-specific adjustment factors for the year are applied to estimates at all levels within the region.

Conclusions

We have applied a standard survey estimation method to derive estimates of traffic from a benchmark sample for minor roads in England and Wales. These estimates are calibrated to road lengths within classes determined by broad road type and region, in order to compensate for minor under-coverage and some sample loss.

The estimates are available broken down by region and by vehicle type. We recommend adjusting existing estimates to the benchmark regionally. This is in part due to large adjustments for some vehicle types that would be required, which may be due to volatility in both sets of estimates for those relatively rare types. Also, the method of estimation for the rolled forward estimates accounts for change by vehicle type but not by region, so there appears to be more of a risk of accumulated error by region.

As the benchmark adjustments are intended to account for accumulated error from an earlier point, we recommend that the adjustments are tapered back to that point to ensure continuity at that point and spread the impact of the adjustments smoothly over years.

Appendix A: Calculation of expansion factors

For each day of the year, *d*, the cleaned ATC count at each ATC count point, *i*, within expansion group, *e*, is divided into day and night counts: $z_{eid} = z_{eid}^{(d)} + z_{eid}^{(n)}$. An *expansion factor* is created for each site (366 used here as 2008 is a leap year):

$$f_{eid} = \bar{z}_{ei} \big/ z_{eid}^{(d)}, \ \bar{z}_{ei} = \sum_{d} z_{eid} \big/ 366$$

So, each of these expands from the daytime flow on any particular day to the average annual all day flow for a site. (This means that when a daytime count is large relative to other days in the year, the expansion factor can be less than one.)

Then the expansion factor used in practice is the median across sites for that expansion factor category and day:

$$f_{ed} = median_{i \in (id)} \{ f_{eid} \}$$

Appendix B: Cell sizes for the calibration

The following table shows the set and counted benchmark sample broken down into the classes used in the road length calibration, together with the counted sample expressed as a percentage of the whole sample, or *count rate*.

Across all cells the count rate is high, 92%-99%, with the notable exception of Wales where only 70% of B roads and 64% of the other minor roads were counted. This is the result of sub-sampling these to reduce the achieved sample to match the available budget. Since this appears to have been carried out at random, the calibration here should counter the impact of that sub-sampling.

As well as the clear distinction between the rates for England and Wales, the classification also captures a systematic difference in the count between B roads and other minor roads with the B-road count rate higher in all regions. London has a lower count rate than average but this can be wholly accounted for by the omission of counts for one borough.

Table 5: Sample sizes for the counted sample and full set sample

		Set sample			Counted sample			Counted sample as a percentage of set sample		
NUTS1		All	B roads	Other minor roads	All	B roads	Other minor roads	All	B roads	Other minor roads
UKC	North East	552	85	467	526	84	442	*95*	*99*	*95*
UKD	North West	1,208	240	968	1,118	224	894	*93*	*93*	*92*
UKE	Yorks. & the Humber	940	183	757	891	176	715	*95*	*96*	*94*
UKF	East Midlands	849	163	686	792	156	636	*93*	*96*	*93*
UKG	West Midlands	1,196	254	942	1,150	248	902	*96*	*98*	*96*
UKH	East of England	1,382	286	1,096	1,290	279	1,011	*93*	*98*	*92*
UKI	London	805	164	641	740	151	589	*92*	*92*	*92*
UKJ	South East	1,890	359	1,531	1,805	351	1,454	*96*	*98*	*95*
UKK	South West	1,067	264	803	1,029	262	767	*96*	*99*	*96*
UKL	Wales	594	135	459	386	94	292	*65*	*70*	*64*
All	England & Wales	10,483	2,133	8,350	9,727	2,025	7,702	*93*	*95*	*92*

Appendix C: Alternative benchmarked estimates

We consider two possible estimators for an overall total at time *t*:

$$\hat{T}_{1t} = (\hat{R}_t / \hat{R}_0)\hat{B} \ \ \& \ \ \hat{T}_{2t} = \sum_c (\hat{R}_{tc} / \hat{R}_{0c})\hat{B}_c$$

In the first of these a rolled-forward estimate $\hat{R}_t$ at time *t*, is adjusted by a factor equal to the benchmark estimate $\hat{B}$ divided by the rolled-forward estimate at the benchmark time $\hat{R}_0$. In the second estimator, the same adjustment is made, but disaggregated by a classification which could be by region, local authority or any other classification.

These look like combined and separate ratio estimators, but with benchmark estimates rather than a fixed total.

Using the approximation for a product:

$$\mathrm{var}(yx) \approx x^2 \,\mathrm{var}(y) + y^2 \,\mathrm{var}(x) + 2yx \,\mathrm{cov}(y, x)$$

then:

$$\mathrm{var}(\hat{T}_{1t}) = \hat{B}^2 \,\mathrm{var}(\hat{R}_t / \hat{R}_0) + (\hat{R}_t / \hat{R}_0)^2 \,\mathrm{var}(\hat{B})$$

and:

$$\mathrm{var}(\hat{T}_{2t}) = \sum_c \{\hat{B}_c^2 \,\mathrm{var}(\hat{R}_{tc} / \hat{R}_{0c}) + (\hat{R}_{tc} / \hat{R}_{0c})^2 \,\mathrm{var}(\hat{B}_c)\}$$

since covariances between benchmark and rolled forward estimates are zero and regional estimates are independent. Note that these variances and covariances treat the sample of ATC counts as fixed and relate to the distributions generated by the sampling of the count points for the original sample and benchmark sample.