Population Synthesis: Comparing the Major Techniques Using a Small, Complete Population of Firms

Justin Ryan, Hanna Maoh and Pavlos Kanaroglou, Center for Spatial Analysis (CSpA), School of Geography and Earth Sciences, McMaster University

Introduction

Recently, disaggregate modeling efforts have received wide attention in the literature (Ballas, Clarke, Dorling, Eyre, Thomas & Rossiter, 2005; Beckman, Baggerly & McKay, 1996; Frick & Axhousen, 2004; Huang & Williamson, 2002; Simpson & Tranmer, 2005; Williams, 2003). There are a variety of reasons for this, including gains in computing power and data availability. As disaggregate models gain popularity, the need for quality micro level input data increases. In general, population micro data is either suppressed to maintain confidentiality, or incomplete due to the high cost of its acquisition (Moeckel, Spiekermann & Wegener, 2003). Consequently, population synthesis techniques are devised as a viable alternative to the collection of micro data, for use in disaggregate models (Beckman, Baggerly & McKay, 1996; Harding, Lloyd, Bill & King, 2004).

Population synthesis techniques are algorithms that take aggregate population data, as well as sample population data as inputs, and produce a complete list of a population's members, each with associated attribute data, as output. Various types of synthetic populations can be created to suit different needs; these may include individual, household, dwelling and firm populations, to name just a few.

Two population synthesis techniques that have emerged as dominant in the literature are the Combinatorial Optimization (CO) technique and the Synthetic Reconstruction (IPFSR) technique (Huang & Williamson, 2002). In this paper, we use both of these techniques, in order to test their ability to recreate a small, complete population of firms for the Hamilton, Ontario CMA in the year 1990. From the complete firm population (11, 499 in total), different levels of input data are extracted. The techniques are implemented with these different levels of input data, and outputs are compared to the entire population, in order to explicitly test their quality. The purpose of this paper is fourfold: to implement the CO and IPFSR techniques for general use; to compare the two techniques, by measuring each one's ability to recreate the known population; to ensure that for both techniques, higher quality input data yields higher quality synthesized populations; to gain an idea of the minimum input data requirements for each technique to produce synthetic populations of reasonable quality. These objectives are realized through a series of comparisons of the outputs from both techniques, using various levels of input data, to the known population.

The remainder of this paper is organized as follows. Section 2 provides a background on the mechanics of the two population synthesis methods used in the paper. Section 3 discusses the methods of analysis adopted to fulfill the objectives of the study. Section 4 presents the empirical results and provides a discussion and a conclusion of our study. Finally, the last section provides some insights on the limitations of the study and discusses further research work on the topic.

Background

The basic ideas behind population synthesis are straightforward. Given whatever population information is available, a list of members of the population is created, using some algorithm, such that the synthetic population conforms to the base information. Then it can be said that the synthetic population is among the set of 'best possible' estimates of the actual population, given the input information. Of course, different algorithms may produce synthetic populations, which all conform to the input data, but differ in their quality. This is the result of each algorithm having its own underlying theory, which may or may not be sound. A common situation is to have input data consisting of a small sample from the population (generally with no

spatial identifiers), as well as tabulations representing the distribution of population characteristics over space. It is to this situation that the CO and IPFSR techniques are particularly well suited.
The Synthetic Reconstruction technique is presented by Wilson & Pownall (1976). Here the emphasis is on the creation of synthetic populations, given a known multiway table of conditional probabilities pertaining to population characteristics. Technical aspects of the technique such as the ordering of conditional probabilities in the selection of attributes are explored fully, and an example of a synthetic population of households is presented. Variations of the Synthetic Reconstruction technique, generally making use of the Iterative Proportional Fitting technique for the creation of multiway tables, are widespread (see for example: Smith et al., 1995; Beckman et al., 1996; Huang & Williamson, 2002; Frick & Axhousen, 2004; Walker, 2004; Simpson et al., 2005; Ballas et al., 2005; Arentze et al., 2007; Guo et al., 2007).

The Combinatorial Optimization technique and its variations are far less common in the literature than Synthetic Reconstruction techniques. One major effort to synthesize populations using these techniques has been undertaken by the National Centre for Social and Economic Modeling (NATSEM), centered at the University of Canberra, Australia (see for example: Harding et al., 2004; Melhuish et al., 2002; Williams, 2001). Another effort at population synthesis, using Combinatorial Optimization, from which the techniques described in this paper are more directly derived, can be found in Voas & Williamson (2000) as well as Huang & Williamson (2002). In the latter paper, the CO and IPFSR methods are directly compared in their abilities to produce synthetic microdata. It is concluded that although both methods produce reliable synthetic microdata, there is less variation amongst populations produced using the CO method, as compared with those created using the IPFSR method. Therefore, the CO method is deemed superior to the IPFSR method.

Briefly, the IPFSR method is as follows. Using the IPF method, multi-way tables of every possible combination of constraining characteristics are created. Constraining characteristics are relevant population characteristics such as Age, Sex or Income. These are usually represented as tabulations or cross tabulations, such as Sex by Census Tract, or Age by Census Tract. So for instance, if 4 constraining characteristics are to be used in the synthesis, multi-way tables of every subset of these 4 characteristics must be provided, or estimated using IPF. The final goal is to obtain the multi-way table of all constraining characteristics. During this process, if an n-dimensional table is being estimated, then the previously estimated n-1 dimensional tables are used as the marginal target values, while the initial state of the n-dimensional table is determined by the population sample. Once the final multi-way table is created, members of the population are created one by one by conducting a series of Monte Carlo simulations using the conditional probabilities defined by the table. See Huang & Williamson (2002) for a more detailed explanation.

The CO method is conceptually much simpler than the IPFSR method. Here, the population for each spatial division of the study area is synthesized in sequence. In particular, for a given area (for instance a census tract), a random sub-set from the sample is selected having the same size as the area's population. This sub-set is assessed for its fit to the constraining tabulations. Following this, one member of the sub-set is randomly switched with another from the sample, and the fit is re-assessed. If the fit improves, the switch is maintained. This process is repeated until a sub-set of the sample having satisfactory fit to the constraining tabulations is obtained. This sub-set becomes the synthetic population for the given area. Again, see Huang & Williamson (2002) for a more detailed explanation.

Methods of Analysis

Programs to execute the CO and IPFSR methods were written in C++ and are collectively called the Synthpop program. For the CO method, tests were conducted to determine the minimum number of iterations for the method to perform in order to reasonably assure convergence. For synthesis of the 1990 Hamilton firm population, 6000 iterations were found to be sufficient. Similarly, for the IPFSR method, the minimum number of iterations to be used as the cut-off point during the IPF portion of the program was determined to be 100. In both cases, the reason for limiting iterations was to find a balance between proper convergence of the techniques and the time required to run them.

The 1990 firm population consisted of 11, 499 firms spread across the 127 Census Tracts comprising the city of Hamilton, Ontario (see Figure 1). The attributes of each firm were: number of employees; census tract (CT); 3-digit Standard Industrial Classification (SIC) code. The values of these attributes, namely number of employees and 3-digit SIC, were recoded to create representative cross-tabulations as they would be available in practice. The values from the number of employees attribute were reclassified into 6 ordinal categories representing discrete employment ranges. The new attribute is referred to as EmpCat. Two reclassification schemes were used to represent the 3-digit SIC. The first divided the 3-digit SIC codes into 14 mutually exclusive categories, and is referred to as SIC-E. The second scheme divided the 3-digit SIC codes into 68 mutually exclusive categories, and is referred to as SIC-2d. The SIC-E and SIC-2d codes are commonly used means of representing firm industrial classifications, with SIC-E codes providing lesser detail than SIC-2d codes. The CT attribute was not reclassified from its original 127 categories. From these attributes over the entire population, the following representative cross-tabulations were derived: SIC-E by CT; SIC-2d by CT; EmpCat by CT; EmpCatXSicE by CT; EmpCatXSic2d by CT. Here EmpCatXSicE is a variable representing all possible combinations of the variables EmpCat and SIC-E, while EmpCatXSic2d is a variable representing all possible combinations of the variables EmpCat and SIC-2d. The EmpCatXSicE and EmpCatXSic2d variables contained 84 and 408 categories, respectively. It is useful to note that although a given firm could theoretically take on any of the 408 categories in EmpCatXSic2d, in the actual population only 335 of these categories are represented. In addition to the tabulations, eight different samples were taken randomly from the firm population, ranging in size from 1% to 100% of the entire population.

As input to the synthesizing process, three sets of tabulations were used, namely: level A – EmpCat by CT and SIC-2d by CT; level B – EmpCatXSicE by CT and SIC-2d by CT; level C – EmpCatXSic2d by CT. Here the detail of the tabulations increases from levels A to B and finally to C. For each combination of input tabulations and sample sizes, two synthetic populations were created using the CO method, and similarly for the IPFSR method. Specifically, 48 populations were synthesized using each method, yielding 96 synthesized populations in total. Both the synthetic populations as well as the actual population were then represented in two-dimensional tables of EmpCatXSic2d by CT. Since there are 127 Census Tracts and 408 EmpCatXSic2d categories, the tables representing the synthetic and actual populations contained 127 × 408 = 51816 cells. Synthetic populations were then compared to the actual population using the Freeman-Tukey statistic, which is defined as follows:

$$FT^2 = 4\sum_i \sum_j \left(\sqrt{S_{ij}} - \sqrt{A_{ij}}\right)^2 \qquad \ldots(3)$$

Where S_{ij} is the ij^{th} cell from the synthesized population table and A_{ij} is the ij^{th} cell from the actual population.

The inspiration for using the Freeman- Tukey statistic comes from Voas & Williamson (2001), which provides a detailed discussion of goodness-of-fit measures for the evaluation of synthetic microdata. The Freeman-Tukey statistic follows a χ^2 distribution, with degrees of freedom equal to one less than the number of cells in the tables being compared, in our case 51816 – 1 = 51815, giving a 5% critical χ^2 value of 52346. An FT^2 value of less than 52346 indicates that the table representing the synthesized population is statistically similar to the table representing the actual population, at a 95% level of confidence. At the same time, an FT^2 value of 0 indicates that the two tables match perfectly, and in general, the closer an FT^2 value is to 0, the better the match between the two tables. It is important to note that since both the synthesized and actual populations are being represented in terms of EmpCatXSic2d by CT (that is to say, at the C level of tabulation detail) the FT^2 statistic is measuring the fit of synthesized populations to a representation of the actual population.

Nonetheless, the representation of actual and synthesized populations in terms of EmpCatXSic2d by CT contains enough detail to acceptably describe firm characteristics, for most purposes. A further advantage of the Freeman-Tukey statistic is that the presence of zeroes in either the synthetic or actual table is not problematic, as is the case with other statistics following the χ^2 distribution (Voas & Williamson, 2001), which allows for a fairly detailed representation of the population such as EmpCatXSic2d by CT, or more detailed still.

In addition to the many synthetic populations created with the CO and IPFSR programs, two random populations were also created. These populations were random in the sense that the firms belonging to each census tract were randomly assigned EmpCatXSic2d categories. However, the number of firms assigned to each CT was consistent with the actual population, mimicking the CO method outputs in this respect. The reason for creating these two random populations was to compare them to the actual population using the FT^2 statistic, and determine the sensitivity of the statistic to an arbitrarily created population. If these random populations proved to be statistically similar to the actual population, the discernment of the FT^2 statistic would be brought into question.

Finally, 50 new populations were synthesized with each method, using the 5% population sample and level A tabulation detail (EmpCat by CT & SIC-2d by CT). Each synthetic population was then compared to the actual population in a similar manner as above. The results of these comparisons were used to assess the variance in the FT^2 values observed from each methods output. Although the choice of input sample size and tabulation detail to be used for the runs was arbitrary, level A tabulations with a 5% sample was chosen due to its similarity to data that could be obtained in practice from publicly available sources.

Discussion and Conclusions

The result of comparisons between the synthetic populations produced using the Combinatorial Optimization method and the actual population can be found in Table 1. Several general trends are immediately evident. First, as the level of tabular detail increases, so does the accuracy of produced populations. At every sample size, populations produced with the tabular level A (CatEmp by CT, SIC-2d by CT) are less accurate than those produced with tabular level B (CatEmpXSicE by CT, SIC-2d by CT). Similarly, populations produced with tabular level B are less accurate than those produced with tabular level C (CatEmpXSic2d by CT). In fact, the best A level population (which makes use of a 100% sample) is less accurate than the worst C level population (using only a 1% sample). This is not a complete surprise, however, since the C level tabulation is equivalent to the actual population representation, while the A level tabulations provide a far less accurate description of the population. With the most detailed tabular input (level C) and the largest sample input (100%), the CO method produces a synthetic population, which is almost identical to the actual population table ($FT^2 = 615$). All of this implies that there is a consistency to the CO method, where higher quality input yields higher quality synthetic populations.

The result of comparisons between the synthetic populations produced using the Synthetic Reconstruction method and the actual population can be found in Table 2. The results are similar to those from the CO method, where for a given input sample size, increasing the tabulation detail increases the accuracy of the synthesized population. Also, for tabulation input levels A and B, an increase in sample size yields an increase in the accuracy of the produced population. For tabulation level C, the synthetic populations are of roughly the same accuracy, regardless of the sample size. This is due to the fact that the IPFSR method reproduces the most detailed tabulation (which for our population is the C level: CatEmpXSic2d by CT) using the sample, making the sample irrelevant when the C level tabulation is given as input.

Although the accuracy of synthetic populations produced with both the CO and IPFSR methods generally increase as the sample size increases, the gains to be had by a unit increase in sample size are not uniform. For the CO method, output accuracy increases drastically when the sample size is increased between 1% and 5%. From this point on however, there is only a gradual, small increase in accuracy to be had from further increases in sample size. These results hold true for all levels of tabulation detail. Therefore, if there is a cost associated with the collection of sample data, we recommend a 5% sample be used as input to the CO method. In the case of the IPFSR method, a slightly different pattern can be observed. Output accuracy increases drastically for sample size increases between 1% and 2.5%, and again for increases between 20% and 50%. For sample size increases between 2.5% and 20%, relatively little gain in accuracy is observed. This result holds true for all levels of tabular detail excepting C which, as discussed earlier, is insensitive to changes in sample size. We conclude that for the IPFSR method, a 2.5% sample input is sufficient where data collection is costly.

Table 4 contains some summary statistics describing the comparison results (sets of Freeman-Tukey statistics) of populations synthesized using the CO and IPFSR methods. Of particular interest is the fact that the minimum FT^2 statistic from the CO outputs is 540, while that number is 7802 for the IPFSR outputs. This shows that given high levels of input information, the CO method is capable of producing more accurate synthetic populations than the IPFSR method. The maximum FT^2 statistics for CO and IPFSR outputs are 33171 and 33956, respectively. This shows that the worst CO outputs are better than the worst IPFSR outputs, and the best IPFSR outputs are worse than the best CO outputs. It is important to note, however, that the 5% Chi-square critical value for the FT^2 statistic used in these comparisons is 52346, meaning that all of the populations synthesized with both methods are statistically similar to the actual population, with 95% confidence. The two random populations which were produced (see Table 3) have FT^2 values that exceed the 5% Chi-square critical value, lending further credit to the results of both the CO and IPFSR programs. Again referring to Table 4, the mean FT^2 statistic values for the CO and IPFSR methods were 14048.56 and 18610.19, respectively, meaning that the CO method produced superior results on average, over all of the different synthetic populations produced with varying levels of input data.

In practice, level A tabulations are more likely to be obtained than the more detailed level B and C tabulations. For instance, publicly available census tabulation data is usually similar in form to the level A tabulations. For this reason, we directly compared the A level CO and IPFSR outputs across all sample sizes. Here, the CO method outperforms the IPFSR method for all sample sizes, except for 50% and 100%. This implies that for a very large input sample size, and poor input tabulation detail, IPFSR provides superior synthetic populations to CO. However, in the more realistic scenario of small sample size, the CO method is recommended over the IPFSR method.

For the two sets of 50 populations produced with an input of tabulation detail A and 5% sample size, summary statistics of the resulting FT^2 values can be found in Table 5. Of particular note, the mean of the CO produced populations is 21538.73, while the mean of the IPFSR populations is 27300.77. Furthermore, the maximum FT^2 value from the CO populations is 21928.09, while the minimum FT^2 value from the IPFSR populations is 26942.28. Thus, CO outputs at this level of input data are consistently closer to the tabular representation of the actual population than their IPFSR counterparts. This result is important, because the level of input data used in these runs is typical of what can be obtained by a researcher in practice. The standard deviation of the CO outputs is larger than that of the IPFSR outputs. In particular, the standard deviation of the CO outputs is 180.05, while that of the IPFSR outputs is 173.95. Of course, the fact that all CO produced populations in the set have significantly lower FT^2 values than the most accurate IPFSR population, outweighs the advantage these latter populations have in terms of standard deviation.

Limitations of the Study and Further Work

Despite the obvious advantages of working with a small population, this limits the scope of the study. The Synthpop programs may perform in comparatively different ways when dealing with larger populations; in particular, the time required for the CO method to run may become much longer compared with the IPFSR method, as synthetic population size increases. (We make the assumption here that the comparative accuracies of small synthetic populations created with CO and IPFSR methods will be maintained in the creation of somewhat larger populations). Another limitation of the study conducted here is the low number of attribute variables synthesized for each member of the created populations. Again, increasing the number of attribute variables in the synthetic populations could yield unexpected results in the quality of the populations produced by either of the methods.

Further work will include the synthesis of larger populations having more attribute variables associated with them. It would be useful to determine the limits of each program, in terms of the time and computing power required to produce extremely large synthetic populations. Also, various methods of comparison between synthetic and actual populations will be explored, including methods of disaggregate or list based comparison.

Another future project will be to link separate populations in a meaningful way. For instance, if disjoint synthetic populations of individuals, households and dwellings exist over a common study area, we aim to link these populations such that no conflicts arise between the population's attributes.

References:

Arentze, T., H. Timmermans & F. Hofman. (2007) "Creating Synthetic Household Populations: Problems and Approach." TRB 2007 Annual Meeting proceedings.

Ballas, D., G. Clarke, D. Dorling, H. Eyre, B. Thomas & D. Rossiter. (2005) "*SimBritain*: A Spatial Microsimulation Approach to Population Dynamics." Population, Space and Place, 11, 13-34.

Beckman, R.J., K.A. Baggerly, M.D. McKay (1996). "Creating Synthetic Baseline Populations." Transportation Research A 30 (6) pp. 415-429.

Frick, M. and K. Axhousen. "Generating Synthetic Populations using IPF and Monte Carlo Techniques: Some New Results."

(2004) Paper presented at the Swiss Transport Research Conference, March 25-26, 2004.

Guo, J. Y. & C. R. Bhat. (2007) "Population Synthesis For Microsimulating Travel Behavior." TRB 2007 Annual Meeting proceedings.

Harding, A., R. Lloyd, A. Bill & A. King (2004). "Assessing Poverty and Inequality at a Detailed Regional Level – New Advances in Spatial Microsimulation." Research Paper No. 2004/26, originally prepared for the UNU-WIDER Conference on Inequality, Poverty and Human Well-Being, 30-31 May 2003, Helsinki.

Huang, Z. and P. Williamson. (2002) "A comparison of synthetic reconstruction and combinatorial optimization approaches to the creation of small-area microdata." Working Paper. Department of Geography, University of Liverpool.

Melhuish, T., M. Blake & S. Day. (2002) "An Evaluation of Synthetic Household Populations for Census Collection Districts Created Using Spatial Microsimulation Techniques." Paper prepared for the 26th Australia & New Zealand Regional Science Association International (ANZRSAI) Annual Conference, Gold Coast, Queensland, Australia, 29 September – 2 October, 2002.

Moeckel, Rolf, K. Spiekermann and M. Wegener. (2003) "Creating a Synthetic Population." Paper presented at 2003 CUPUM, Sendai, Japan.

Norman, Paul (1999). "Putting IPF on the Researcher's Desk." Working Paper, School of Geography, University of Leeds, United Kingdom.

Simpson, Ludi and Mark Tranmer. (2005) "Combining Sample and Census Data in Small Area Estimates: Iterative Proportional Fitting with Standard Software." The Professional Geographer, 57(2), pp.222-234

Smith, L., R. Beckman, K. Baggerly, D. Anson, M. Williams. (1995) TRANSIMS: Project Summary and Status May 1995. Los Alamos National Laboratory, Report prepared for U.S. Department of Transportation and U.S. Environmental Protection Agency.

Voas, D. and P. Williamson (2000). "An evaluation of the combinatorial optimization approach to the creation of synthetic microdata." International Journal of Population Geography 6 (6) pp.349-366

Voas, D. & P. Williamson (2001). "Evaluating Goodness-of-Fit Measures for Synthetic Microdata." Geographical & Environmental Modeling, Vol. 5, No. 2, 2001, 177-200.

Walker, Joan L. (2004). "Making Household Microsimulation of Travel and Activities Accessible to Planners." Paper presented at TRB 2005 Annual Meeting.

Williams, Paul. "Using Microsimulation to Create Synthetic Small-Area Estimates from Australia's 2001 Census." (NATSEM related paper, 2003)

Wilson, A.G. and Pownall, C.E. (1976). "A new representation of the urban system for modeling and for the study of micro-level interdependence." Area 8 pp.246-254.

Wong, David. (1992) "The Reliability of Using the Iterative Proportional Fitting Procedure." Professional Geographer, 44(3), pp.340-348.

Figure 1: The 11, 499 firm locations in the Hamilton CMA, 1990.

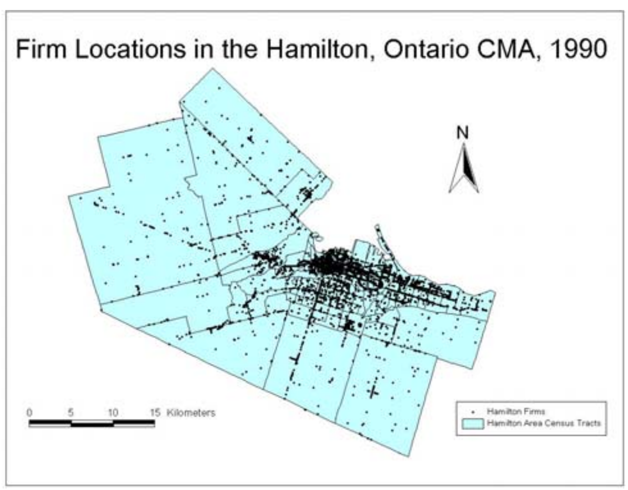

Table 1: results of comparisons between CO produced synthetic populations and the actual population, using the Freeman- Tukey statistic (critical value 52346).

Run #	Sample Size	(A) FT^2 for: EmpCat by CT; SIC-2d by CT	(B) FT^2 for: EmpCatXSicE by CT; SIC-2d by CT	(C) FT^2 for: EmpCatXSic2d by CT	Does the synthetic population fit the actual at a 5% critical chi-square value of 52346?
1	1%	33156	27288	18965	Y, Y, Y
2	1%	33171	27535	18966	Y, Y, Y
3	2.5%	23541	17488	9381	Y, Y, Y
4	2.5%	24048	17718	9381	Y, Y, Y
5	5%	22037	13440	5437	Y, Y, Y
6	5%	21806	12987	5440	Y, Y, Y
7	7.5%	21744	12443	3722	Y, Y, Y
8	7.5%	21304	12475	3709	Y, Y, Y
9	10%	20889	11751	2676	Y, Y, Y
10	10%	20889	11520	2694	Y, Y, Y
11	20%	20561	11012	1491	Y, Y, Y
12	20%	20563	10609	1525	Y, Y, Y
13	50%	19481	10231	568	Y, Y, Y
14	50%	19637	9929	698	Y, Y, Y
15	100%	19291	9899	540	Y, Y, Y
16	100%	19897	10183	615	Y, Y, Y

Table 2: results of comparisons between IPFSR produced synthetic populations and the actual population, using the Freeman- Tukey statistic (cutoff value 52346).

Run #	**Sample Size**	**(A) FT^2 for: EmpCat by CT; SIC-2d by CT**	**(B) FT^2 for: EmpCatXSicE by CT; SIC-2d by CT**	**(C) FT^2 for: Emp CatX Sic2d by CT**	**Does the synthetic population fit the actual at a 5% critical chi-square value of 52346?**
1	1%	33956	26903	7869	Y, Y, Y
2	1%	33831	27221	8201	Y, Y, Y
3	2.5%	27332	21180	7707	Y, Y, Y
4	2.5%	27332	20909	8048	Y, Y, Y
5	5%	27541	19409	7921	Y, Y, Y
6	5%	27093	18675	8119	Y, Y, Y
7	7.5%	27776	18605	7837	Y, Y, Y
8	7.5%	27555	18438	8032	Y, Y, Y
9	10%	26688	17956	8139	Y, Y, Y
10	10%	26559	18084	7837	Y, Y, Y
11	20%	23670	16454	8071	Y, Y, Y
12	20%	23592	16546	7828	Y, Y, Y
13	50%	8131	8217	7940	Y, Y, Y
14	50%	8126	7995	7929	Y, Y, Y
15	100%	7802	8036	8162	Y, Y, Y
16	100%	8097	8085	7941	Y, Y, Y

Table 3: Freeman- Tukey results from randomly synthesized populations.

Run	FT^2	5% Critical Chi^2	Fits? (Y/N)
R1	70769	52346	N
R2	70916	52346	N

Table 4: Freeman- Tukey results from CO and IPFSR outputs, except the 50 run sets.

Summary Statistic	CO results	IPFSR results
Max	33171	33956
Min	540	7707
Mean	14048.56	15945.31
Sample Variance	78562751	77423594

Table 5: Freeman- Tukey results from IPFSR and CO 50 run sets.

Summary Statistic	CO	IPFSR
Mean	21538.73	27300.77
Median	21553.04	27303.82
Standard Deviation	180.05	173.95
Sample Variance	32418.86	30259.89
Kurtosis	-0.6049	0.0834
Skewness	-0.1791	0.0271
Range	773.58	816.49
Minimum	21154.51	26942.28
Maximum	21928.09	27758.76
Sum	1076936.65	1365038.44
Count	50	50