APPENDIX B

The MIT Mass Media Computer Simulation Methodology

The MIT computer simulation methodology was developed under contract from the United States Information Agency to study the sociology of Soviet audiences for mass communications. The MIT team was headed by Dr. Ithiel de Sola Pool, Professor of Political Science at MIT. Its findings were published in a series of reports in 1975, all of which are noted in the select bibliography to this paper. The input data for this initial study came from SAAOR interviews conducted in 1970–1972. Professor Pool gives a detailed account of the methodology used in this first simulation application in the methodological report issued as "Simulation Report 4" in the report series.

The MIT simulation methodology is described in some detail in an article in *Communications Research: An International Quarterly* of October 1982. The article was authored by Professor Pool, Dr. John Klensin, Principal Research Scientist at MIT and R. Eugene Parta, Director of SAAOR. An earlier, detailed example of the application of the simulation methodology was included in "Listening to Radio Liberty in the USSR, 1976–77," Analysis report 3-78, SAAOR, by R. Eugene Parta. Those wishing more details on the methodology are referred to the above publications as space limitations here do not permit more than a highly simplified general description.

The simulation methodology was developed to address the basic issue of how to draw estimates from uneven samples, given that the sample deficiencies could not be corrected in the field.

After applying extensive cross-checks to ensure that the data were internally consistent, the next step was how to estimate underlying data from aggregated results, in this case estimating individual cell values in a contingency table (which in some instances might be quite weak or even missing) from the table's marginals (which would be considerably more robust). The algorithm for this computational process was dubbed "Mostellerization." It had been developed by Prof. Frederick Mosteller of Harvard University (see select bibliography) and was first applied in the simulation process to create a population model of the USSR. The table was based on input from Soviet census data, but the data were not available in a multidimensional format that could be used by the simulation. It was necessary to create a new 4-way demographic table (age by education by gender by rural/urban residence) from lower dimension tables in the census data. This provided a 24-cell table with 3 levels of age, 2 of education, 2 of gender and 2 of rural/urban residence. The next step was to create a 240-cell table by factoring in ten geographic regions. The final step in creating the population model was to expand it to 480 cells by factoring in Communist Party membership, data which was not included in the census but had to be found elsewhere—in this case, in *Partinaya Zhizn'* ("Party Life").

Since samples were insufficient to compute listener estimates in all 480 cells of the population model directly, the second stage of the simulation relied on the Mostellerization algorithm to make estimates of underlying cell data from aggregated listenership figures. This procedure followed the same process as above, but substituted listening ratios for population figures to compute first the 48-cell table, and then expand this to 480 cells to take into account geographic dispersion. The ratings calculated in each cell, multiplied by the population model values for the cell, provided the basic estimate of the audience.

As normal confidence interval tests could not be applied, a special algorithm was developed that provided estimates of cell reliability on a cell-by-cell basis. Estimates of cell reliability are largely dependent on the sample size collected for that cell, and more importantly, the degree to which that cell needed to be weighted relative to other cells. These reliability estimates were made by assessing the impact, on a cell-by-cell basis, that misclassification of a single respondent would have on the audience

estimate. This means that listening tallied in a cell with a low sample population but a high real-world population (e.g., old, uneducated, rural women) would result in a low sense of confidence about that cell. In estimating a confidence range for an entire table this cell would have a strong impact. Conversely, this procedure gives more confidence in cells where the sample population is large and considerable listening is recorded (e.g., educated, middle-aged urban males). In terms of final estimates the direction of bias in the raw sample data coincided with those strata of the population most likely to engage in Western radio listening, the activity being measured. The result is a relatively strong sample in the cells that contribute most heavily to the listening estimates.

In 1986 the "core audience" concept was introduced. This derived estimates precisely from that part of the sample which was most robust and where most listening to Western radio occurred—the urban, educated, adult population—which made up about a quarter of the entire adult population. This permitted listening trends to be charted more accurately than using the entire adult population where confidence ranges were considerably larger.

In the 1980s, the simulation methodology was adapted more specifically to SAAOR's particular needs. When it was determined that Communist Party membership was not a predictor of listening to Western radio, this dimension could be collapsed out of the simulation process, yielding two basic input tables: a 24-cell demographic table (age, sex, education, rural/urban) and a 20-cell geographic table (geographic region by education) which was "mostellerized" into a final 240-cell output table.

Two MIT methodologists deeply committed to the 1980s effort to refine the MIT simulation process for SAAOR purposes were Dr. John Klensin, Principal Research Scientist at MIT and Dr. Ree Dawson, a Harvard-trained statistician. In 1986 Dr. Dawson wrote a paper "Developing a Methodology for Projecting the Audience to Foreign Broadcasts in the Soviet Union" (see select bibliography for details), which proposed a survey-ratio estimator model to build on the MIT simulation process for purposes of analyzing trends in listening. This was applied to the analysis of listening trends through the end of the project in 1990. In 1988 Dr. Dawson developed a log-linear method for imputing sample values for SAAOR estimates of geographic audiences

(see reference in select bibliography) which permitted the geographic estimates to be calibrated more precisely than hitherto. These are the estimates used in the 1988–1989 charts provided in this paper.

The MIT simulation process could be applied to other purposes than estimating audiences to Western radios. When sample sizes were sufficient, SAAOR used the procedure to study the *samizdat* phenomenon in the USSR, and to study overall media consumption patterns, as reported above in this paper.

SAAOR acknowledges its deep debt to Prof. Ithiel de Sola Pool and his MIT colleagues in developing and applying a pioneering statistical methodology that made possible a deeper understanding of the role and impact of Western broadcasting to the USSR during the Cold War.