DELIVERING ISSUE AUDIENCES

Mario Paic, Andrey Ponomarev, Ipsos Connect
Stef Peeters, Michael Debels, CIM

Introduction

For over 40 years, the CIM (Center for Information about Media, the Belgian JIC) has been measuring and reporting the reading of print titles for the purpose of media planning for the advertising market.

During that time, faced with important changes in press distribution and consumption, the CIM has made a number of essential methodological changes – most notably in 2012 when the questionnaire was redesigned in order to start capturing reading of digital version of print titles. This led to the publication of two new reach indicators besides the Average Issue Readership of the paper version:

- "Paper + Digital Versions - The audience of the paper version as well as the digital editions available in PDF or in tablet and smartphone applications.

- "Total Brand" - The audience of the paper version, the digital versions (PDF + App) and the website of the title.

The 'CIM Press 2016-2019' study marks another important milestone in the continuing evolution of measuring the news and magazine brand audience in Belgium. With this new project, CIM wants to continue measuring and delivering currency audience data using high quality methods and at the same time change the scope of the press audience study to reflect better the dynamics of the press sector by introducing much more granularity into the currency data.

The tendering process for the new CIM Press Study has been conducted during late 2015 and early 2016. It was structured around 7 separate modules covering all aspects of fieldwork and data processing with a separate tender for a new reporting solution following soon after. Ipsos Connect has been awarded with the contract for developing and delivering solution for all 7 core modules - Ipsos teams in Belgium being responsible for the 3 fieldwork modules and the Ipsos Connect Data Science team based in London, UK responsible for developing solutions for the 4 data processing modules.

This paper will focus on the three data processing modules:

1. Module 4 – Creating dated daily readership estimates for newspapers

2. Module 5 – Creating specific issue readership estimates for magazines

3. Module 6 – Modelling magazine readership accumulation

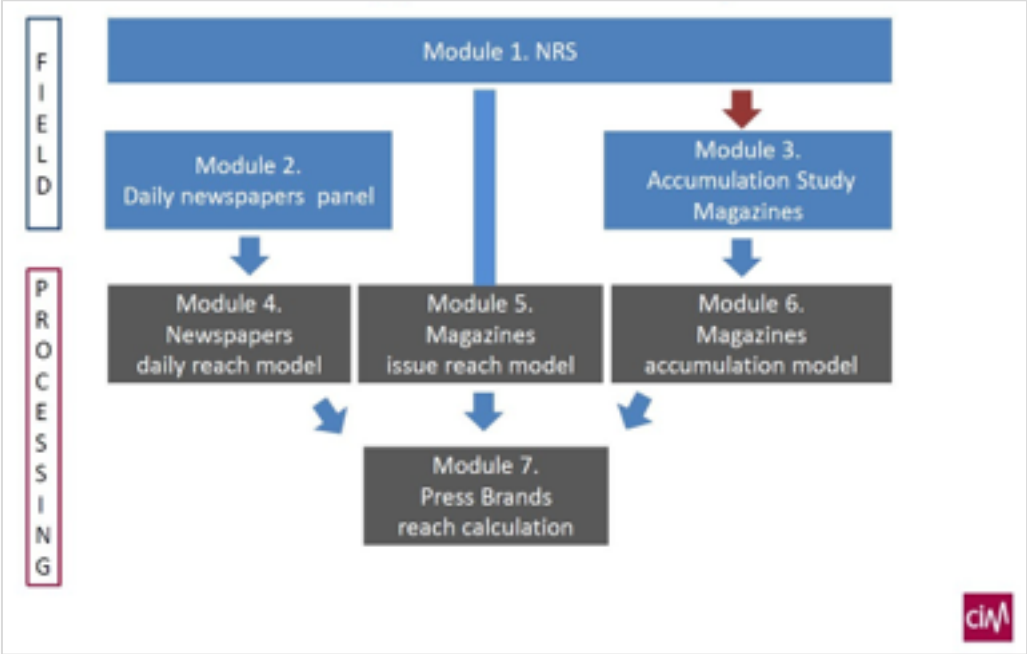
We will describe the general approach and the individual technical procedures used for each of the modules, present some of the challenges that we were faced with while developing the solutions, and offer some general conclusions and recommendations on the implementation of each of the solutions into the standard currency systems - all based on the tests done during the development phase. The paper will offer only basic information on the fieldwork modules and will not cover the data fusion module (Module 7 – Creating brand audiences).

At the time of writing this paper, the first new data from the CIM Press Study 2016-2019 was yet to be released so the paper will not include any official data and will not make any conclusions on the acceptance and real use of the data by the market.

1. Overall design of the new study

As mentioned in the introduction, the overall design of the new CIM Press Study 2016-2019 project includes seven separate data collection and data processing modules, as shown in the diagram below.

Figure1: The overall design of the CIM Press Study 2016 – 2019.



The core component, the main NRS study, remained largely unchanged – nationally representative survey aimed at capturing currency AIR data using standard methodology, including a high-quality "face-to-face" recruitment and a proven CASI/ CAPI interviewing.

The two new components, Modules 2 and 3, were aimed to collect more granular readership data that would underpin the new solution for reporting day based audiences for newspapers and issue based audiences for magazines with an additional function of reporting audience accumulation for magazines.

Different methods were proposed and considered to collect such data. After the review process, two relatively new, but already proven methods, were chosen by CIM:

1. A separate daily SMS panel to collect dated daily readership.

2. Specific issue readership questions based on cover recognition method, included in the main NRS survey, to collect issue-by-issue readership and accumulation for magazines.

The daily SMS approach that was implemented in the new CIM Press Study was based on the existing Ipsos Dailymetrie project, developed in 2013 for one of the newspaper publishers in Belgium. This unique methodology proved to be very successful in transforming a commodity such as traditional text messages, into a robust and reliable data collection platform fit for the specific purpose of capturing information on daily variations in newspaper readership[1].

On the other hand, the idea of using cover recognition to capture magazine readership has been around for some time and it has been a part of the official currency data in two countries – in the Netherlands (provided by GfK for NOM) and Australia (the "emma – Enhanced Media Metrics Australia" survey conducted by Ipsos for The Readership Works).

[1] Van der Steichel A. (PDRF 2015): "Understanding cross-platform readership on a daily basis"

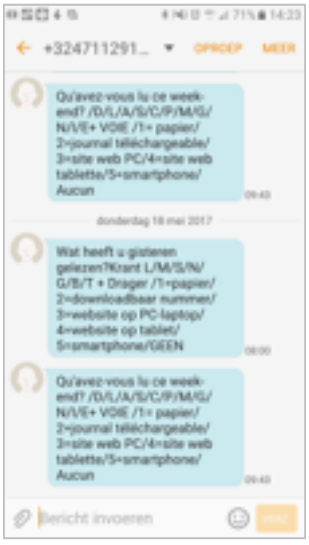
In both markets, the method is primarily used for estimating Average Issue Readership by obtaining the average claims made for the oldest copy shown.

There are several advantages with this approach, including arguably better participant stimuli, elimination of replication errors and the generation of readership accumulation data. This last advantage is particularly important. It is well known that the published readership of a magazine is not instantly created during the week or month of issue. In fact, it can take several weeks or months for the readership to build or accumulate.

CIM decided to employ this method and add cover recognition based SIR questions into the main NRS survey. CIM chose not to use cover data to estimate AIR but to provide continuity by using the standard "Recent Reading" question for estimating AIR. Cover recognition data will be used to underpin the new SIR model. Below images are showing examples of these two methods used.

Figure 2: SIR question based on a cover recognition method as implemented in the new CIM Press Study.

Figure 3: An example of a daily newspaper readership question sent to the CIM SMS panel.



2. Module 4 – Modelling dated audiences for newspapers

Introduction

As explained in the previous section, one of the new components introduced by the new CIM Press Study, is the 3,000 strong daily SMS panel used for collecting data on "day-by-day" readership of newspapers.

It's important to emphasise that this data alone, could be used to provide daily readership estimates. However, following on the work by Frankel et al., described in their 2007 PDRF paper called *"Issue Specific Estimation – Mathematical and Statistical Issues, Procedures and Models"*, our strong belief is that by combining the survey based estimates with the ones created using a bespoke readership model, the final estimate can be significantly improved.

To illustrate why we think the combined approach is needed, we can use a very simple example to outline the limitations of the survey only approach. For example, if a newspaper has an all adults AIR of 10% in the Dutch-speaking region of the country, the error range (at the 95% confidence level) is +/- 1.5 points or 15% of one day's audience, based on a sample of 1,500.

The goal of the modelling procedure is to smooth the sampling error component of the variations in dated audience estimates. Given the fact that we can calculate standard error for both the model and the survey based estimates, we can then merge these estimates to obtain final daily readership figures for delivery to the market.

Method

In this section, we will cover the approach for creating a true "day-by-day" readership estimate for newspapers, outlining the following processes:

a) Processing SMS panel data

b) Creating a dated readership model

c) Merging model and survey estimates to obtain final dated readership estimates

d) Integrating data obtained in 1-3) to the main NRS database

On the first step, the daily SMS panel data is processed:

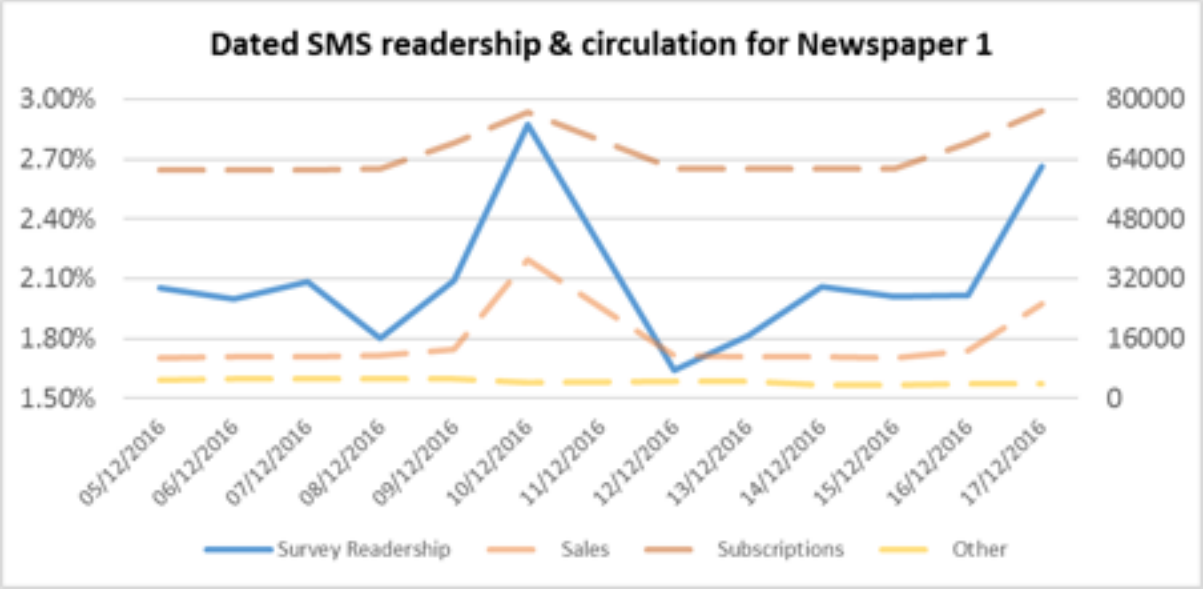
1. SMS panel data is filtered to the valid respondents on a given date
2. Data is weighted, according to population targets
3. Based on the weighted data, **survey based** estimates of dated newspaper readership are obtained

As already mentioned, in creating the dated readership model for the new CIM Press Study, we have been inspired by the work done by Frankel et al. who in their 2007 paper suggested an effective method for smoothing survey estimates by smoothing estimates to the mean. We have tried to accommodate, but also improve, this approach for this particular project. Our basic idea was that this approach may be improved further by smoothing to more than just average. Indeed, there are various predictors of readership which can be utilized to build a robust predictive model, predicting dated readership and then smoothing survey estimates to that, improving the final result.

Even though the topic of readership and circulation and correlation is controversial (see, for example, http://www.newsmediaworks.com.au/the-relationship-between-circulation-and-readership), we still believe that circulation is one of the strongest predictors of readership. As such, the models we currently utilize are based almost uniquely on that. We do, however, expect to include such factors as seasonality, holidays, etc. as we collect more information over time.

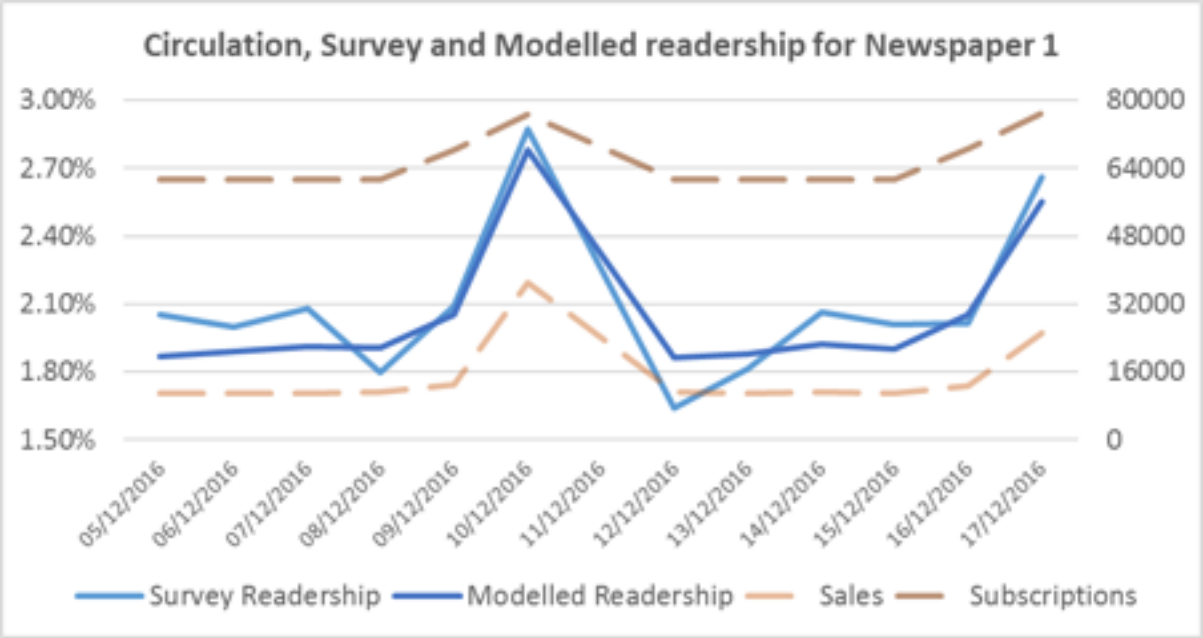
On the chart below, you can see an example of readership and circulation dated trends for one of the newspapers in Belgium. One can see that readership follows circulation fairly closely, especially so with increases on Friday and Saturday.

Figure 4: Comparison of trended dated SMS readership and circulation



SMS panel readership data and circulation data then become a basis for the model, aimed at predicting readership based on circulation. Using the example of the same newspaper, the chart below shows comparison between SMS panel based readership, circulation and modelled readership. One can see that the model is not simply an "average weekday" – changes in circulation affect modelled readership to a great extent. For example, predicted levels for Saturdays of 17/12 and 10/12 are visibly different.

Figure 5: Comparison of dated SMS panel readership, circulation and the model based readership



Once we have **modelled** and **survey** readership, we need to find a way to combine these estimates to obtain an estimate closest to the real readership. Both modelled and survey estimates have their pros and cons:

1) **Survey** readership measures readership directly, but on a relatively small sample. As such, it is an estimate with low bias, but a significant variance.

2) **Modelled** readership estimate is relatively stable, as it is based on circulation data, but does not represent all the true changes in readership. So it is an estimate with higher bias but low variance.

Our goal is to combine these two estimates to get one which would be optimal – final (or "smoothed") estimate. Essentially, the task is to find such proportion of survey and modelled readership where expectation of correlation between final and true readership is the highest:

As tempting as it was for us to use the same procedure as Frankel et al. (described in the 2007 PDRF paper), we were not able to follow the exactly same procedure. In this study, the samples for particular issues of a title were different, hence observations were independent. In the case of the dated newspaper readership data, our daily observations are strongly correlated, due to the panel nature of the underlying survey data.

Our research has shown that there are no analytical methods which would allow us to overcome this issue and merge the estimates using a "formula". In this case, a common practice is to resort to running simulations. We have set up our stimulations in the following fashion:

1) We created a "synthetic population", with the following rules:

    a. Aggregated readership follows our modelled readership estimate.

    b. Frequency distribution is the same as on the SMS panel.

    c. Demographics are representative of population.

2) We then created 1,000 synthetic populations based on step 1, with extra noise added to readership, so that it is different from modelled estimates.

3) For each of these "synthetic populations", we have performed the following procedure:

    a. Create 1,000 samples, same as our SMS panel.

    b. Process this samples the same way we process SMS panel.

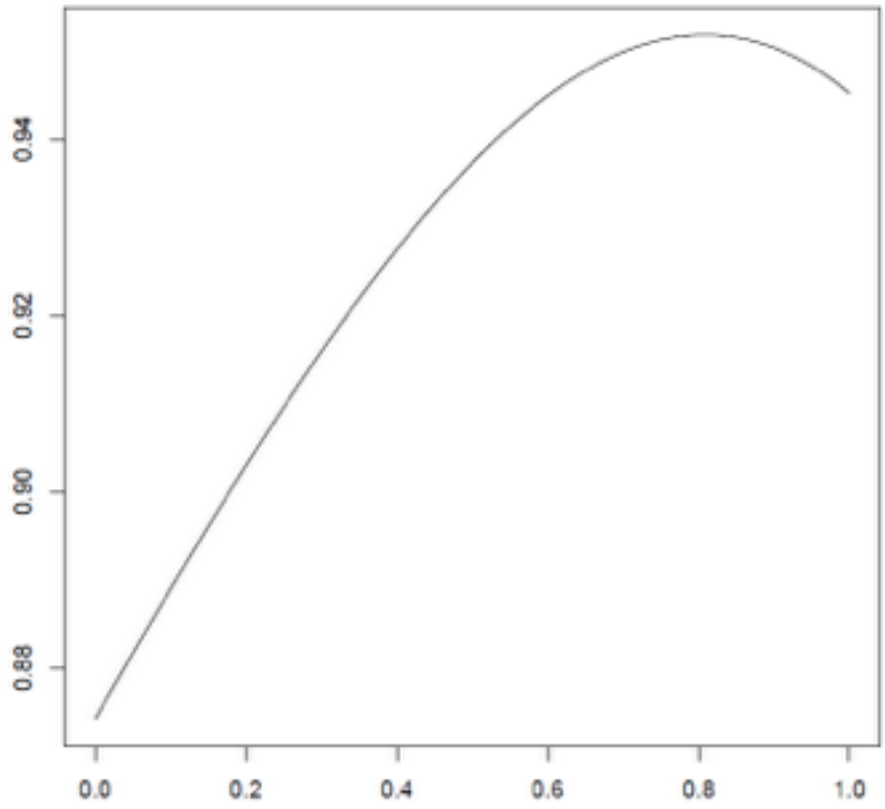
    c. Create a dated readership model on these samples.

4) From the thousands of simulations obtained at the step 3, we have chosen ones closest to our "real-world" result in terms of readership model quality.

5) Based on simulations chosen at the step 4, we calculated the ratio between modelled and survey readership which creates an estimate best correlated to the "population" readership for this particular simulation.

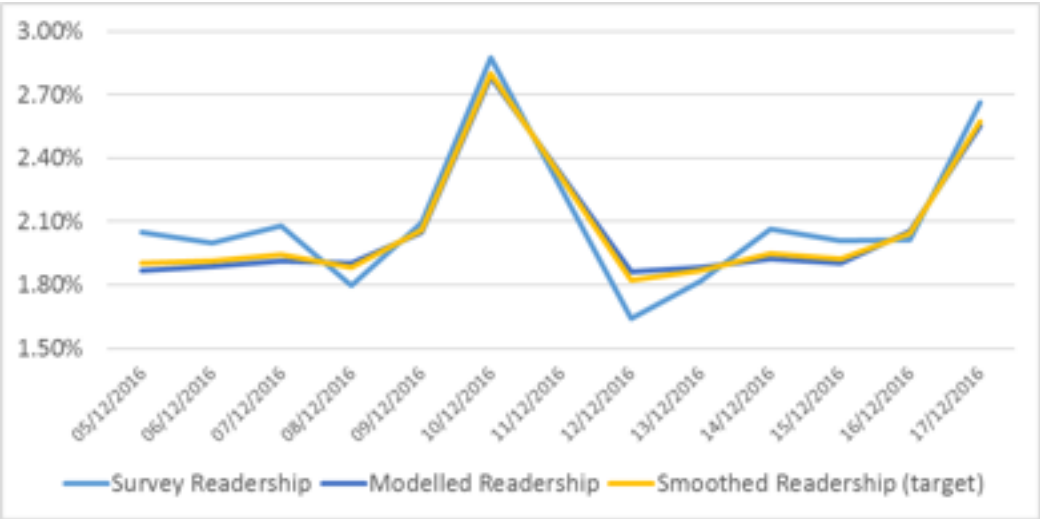
The chart below shows the dependency of correlation between final and “true” readership on synthetic populations for Newspaper 1. One can see that going with pure survey or pure modelled estimate is strictly worse than a blend of the two.

Figure 6: Correlation between the final and “true” readership



This procedure has allowed us to obtain for each of the newspapers, and thus calculate an optimal way to merge modelled and survey estimates to obtain final (or “smoothed”) readership estimate.

Figure 7: Comparison of survey, modelled and the final “smoothed” readership for Newspaper 1



As the last step, before integration into respondent level data, the final merged ("smoothed") readership is converted to an index vs. the average and these indexes are then applied to the main survey Average Issue Readership in order to bring the dated daily readership into the currency data.

Challenges and conclusions

One of the interesting challenges we have faced as the part of the process was how to integrate the obtained daily readership estimates while giving users the opportunity to perform post-campaign evaluation on resulting data.

Traditionally, print media planning is based on AIR probabilities. We could potentially adjust R&F model to accommodate daily AIRs, but it would most likely not be an elegant correction.

On the other hand, for other media, most notably TV, media planning and R&F calculations are done by "head-counting" people in the database, since these studies are traditionally panel-based and do not require reach expansion.

Given that the new CIM Press Study uses a SMS panel for collecting daily readership data, in our integration approach, we have tried to produce a fully integrated dataset that would by definition allow R&F calculations to be done by "head-counting" without any model based calculations needed to be done within the reporting software.

To be able to do that, we first had to perform a data fusion between the main NRS survey dataset and the SMS daily readership dataset. The resulting dataset, preserves all the respondent records from the main NRS survey, but is expanded with raw daily readership data stored as a binary flag for each day during the 12-month period the main NRS database corresponds to. This raw readership data is then calibrated using a custom imputation procedure in order to match the final daily readership targets obtained through the process of merging survey and model based readership described above.

Keeping in mind the fact that the first official results have yet to be released, and that there's still ongoing work on fine-tuning the first version of the model, we can conclude that our basic hypothesis, that the pure survey (SMS panel) based estimate of daily readership can be further improved by merging it with modelled one, has proven to be true and that we can safely recommend for this approach to be used in a regular production of dated daily readership data.

3. Module 5 – Modelling issue audiences for magazines

Introduction

Similar to creating dated daily readership of newspapers, we used, to create issue based audience for magazines, a modelling solution that combines rolling specific issue readership data (collected using cover recognition method), circulation data and the overall AIR levels measured on the main readership study.

The cover recognition method has been tested by CIM in the past with the primary goal to assess if the method can be used to replace the traditional method of collecting data for estimating the Average Issue Readership. The test has produced mixed results and the overall recommendation was to not use the method for estimating AIR[2].

However, the method has been revisited as a potential solution for the Module 2 of the new CIM Press Study which was designed to capture data on reading specific issues of magazines. It was recommended by Ipsos and chosen by CIM as the most appropriate method to collect survey based data on reading of specific magazine issues.

It is our view that, in order to report issue-specific readership, results should be underpinned by survey based estimates of individual issue reading, as well as by circulation data. Even if not used as currency numbers themselves, SIR data provide information invaluable for generating final audience estimates.

Having a rolling sample for each of the issues, with 4 issues being simultaneously surveyed, ensures that we have a large enough sample for estimating every kind of periodicity in the end analysis. For example, if four covers are asked for a weekly title it means that each is asked about for four weeks in the survey. This means basing estimates on around 800 responses every week (assuming an annual sample of 10,000). For a monthly title, sample by cover would reach more than 3,000, creating a strong basis for readership estimation.

Additionally, survey data on the number of issues read creates a strong basis for estimating readership probability at a respondent level, effectively replacing (or enhancing) data from the frequency question.

However, despite the richness of the SIR data, and stable sample sizes, our approach has been enhanced by the use of circulation based modelling and AIR corrections in order to correct for two biggest short failings of solely using a specific issue readership approach based on cover recognition:

1. Firstly, cover recognition suffers from memory decay in the situations where significant time has passed in relation to the publication of the issue – which we aimed to correct through usage of AIR to scale the final levels.

2. Secondly, due to the estimate being sample based, it suffers from sample bias and sample error – which, concerning relatively low samples of actual readers for each issue, would mean that a lot of observed variation is a product of sample error and not the actual variance – this was alleviated by using circulations to produce a more stable model of readership for each title.

3. Finally, due to the samples being controlled at an overall level, and not period-by-period, we had to control for differences in sample structure for each issue being estimated, to ensure that sample skew doesn't produce variations that are not a true measure of readership variation.

[2] Peeters S., Schmerber J.S., Eeckhout L, Cools B (PDRF 2011): "Can Cover Recognition Replace Recent Reading"

Method

To reach the goal of a stable modelled SIR measure, we employed multiple techniques over several stages of production.

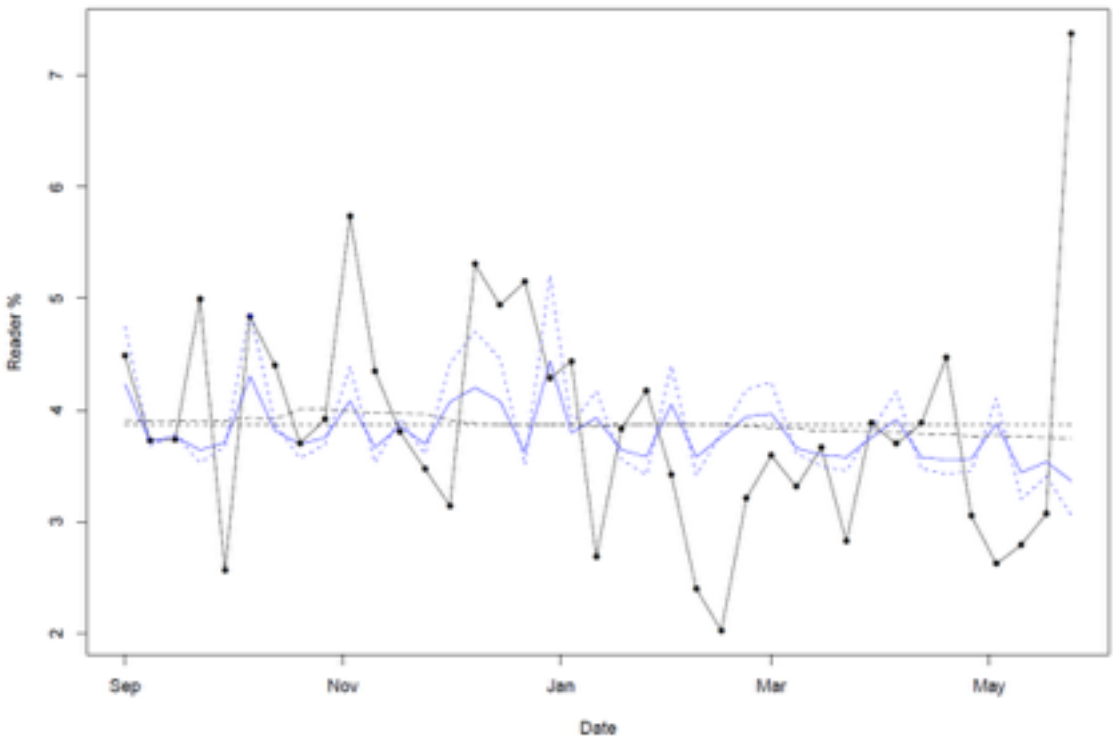
We start with the analysis of survey data – the survey sample is split, for each title, into "issue-by-issue" samples. At the level of the total sample, a CHAID tree procedure is ran to determine which demographic variables are significant drivers of readership for each title. Once we determine them, we record what the readership structure is, and weight each issue sample to the average structure, in order to account for sample skew from issue to issue. This ensures that the starting estimate is based on the same sample assumptions – meaning we can compare them to the AIR in the same way.

The next stage in the process is to produce a SIR estimate that's scaled to the AIR levels in the main NRS study. We do this by averaging issues over periods, to get the index of a periods' estimate to the AIR estimate. This index is then applied to each individual issue and an average across periods is taken to get the resultant estimate of readership for a specific issue.

k– number of issues
m – number of periods
R_{ij} *– readership of issue j for period i*
IR_i *– Index of readership for period i*
AIR – Average issue readership
R_j *– Readership of issue j*

Once we have obtained the specific issue readership estimates, we combine them with the circulation data in order to model an estimate of actual readership. To achieve that, we applied a regularized general linear model, using sales, subscription and other distribution data as predictors, controlling for sample size variance, and using the adjusted specific issue readership estimates as the dependent variable.

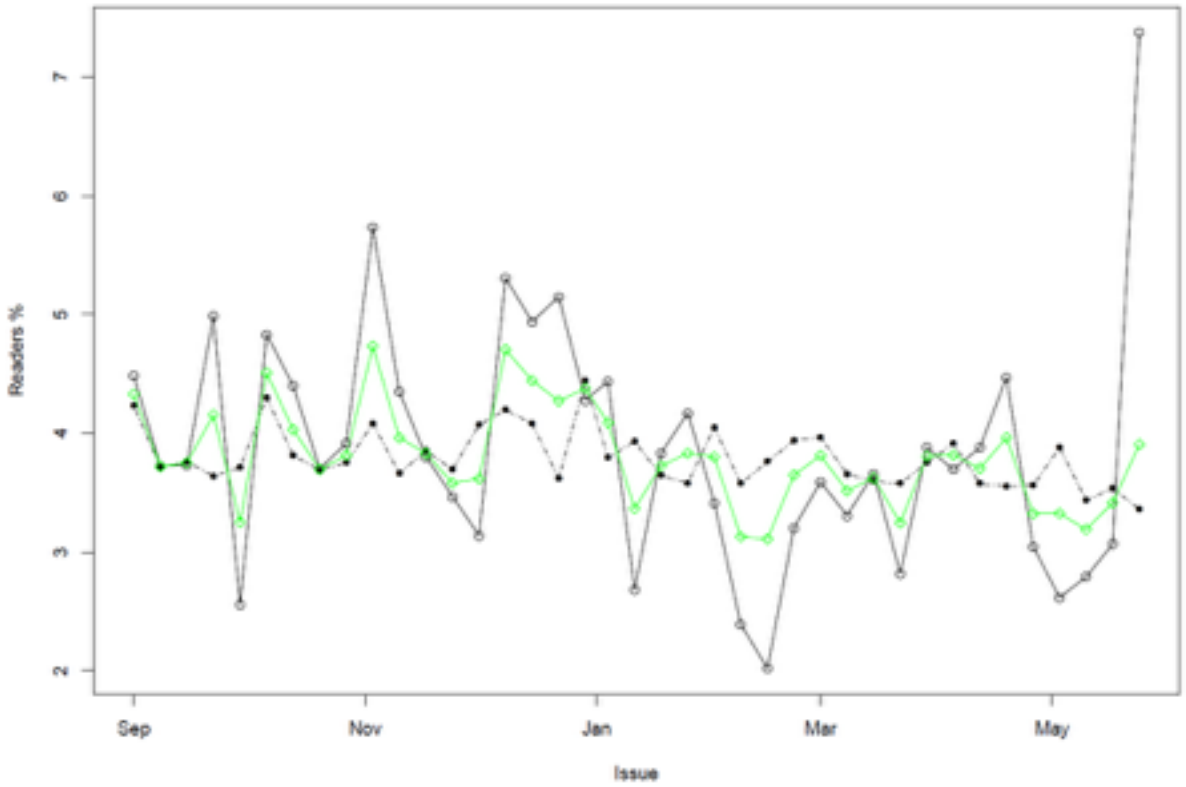
Figure 8: Readership (black) and modelled readership (blue) with circulation modes (dotted) for a title example



Finally, similarly to the procedure used for dated newspaper readership, in order not to lose all the variation present in the survey estimates, we decided to merge the model estimate and the survey estimate using a similar technique based on the work by Frankel et al. The procedure was modified to an extent so that when applied, it did not use mean readership and it treated the modelled estimate as the "ground truth".

It's worth to note that, due to the nature of the survey, issues towards the beginning and end of the surveying period necessarily end up with significantly lower samples, and as such, it is decided that issues that are too close to the end of the surveying period wouldn't be reported for until the next update. We are including them in figure 2, however, to show how sample sizes influence the size of the heteroscedasticity correction.

Figure 9: Readership (black) and modelled readership (dotted) with the resulting merged readership estimate (green)



Challenges

Although the procedure was tested on historic data beforehand, we were aware that the solution is mostly going to rely on good quality circulation data being delivered. We also assumed minimal sample variations week-to-week, and fairly stable publication schedules.

The first issue we've encountered is large sample variability from issue-to-issue due to unstable publications schedules, which made any approach relying on an average sample unviable, meaning that we had to rework the whole procedure with sample size controls being implemented. To address this problem to some extent, we split each issue into weeks as the minimal periods, instead of using expected publications period for each issue. This greatly remedied the problem, but meant that we would be using reduced sample at each "period" point – increasing the number of cases observed for monthlies from 4 to 16. Fortunately, due to using averaged and rolling assessment during each step of the procedure, this was alleviated, but allowed us better control over artefacts arising from memory decay.

There was still the issue of some issues jumping ahead of their publication schedules, due to special issues, etc. (e.g. a weekly issue publishing a new, special issue, just 2 days after the publication of the regular weekly issue). This resulted in samples across different "age" patterns being irregular, and given that nothing could be done from a data analysis perspective, a decision was made to treat those titles separately in the questionnaire design moving forward.

The second issue was related to the state of the circulation data being delivered. Although we received circulation data for each title in the sample, for many titles we were faced with two challenges:

1. Firstly, for a significant number of titles, circulation figures proved to be rather "flat" and show little to no variability. This prevented us from applying viable circulation based models for those titles.

2. Secondly, for some titles, circulation surges preceded readership data. This suggest that lagged circulation might be used as an additional predictor.

A final challenge was the fact that the limited sample sizes for very small titles, combined with relatively "flat" circulation data, were the reason why for a considerable proportion of titles viable models could not be created. In those cases, again by following recommendations from the work done by Frankel et al, readership smoothing towards the mean was performed, in order to reduce the amount of variation estimated by the survey.

Conclusions

Despite the challenges in this particular survey methodology when applied to irregular publication schedules, and some questionable circulation patterns, based on our initial tests, we can conclude that from a technical point of view, the circulation corrected SIR model is deemed as good and reliable method of estimating specific issue readership for medium to large magazine publications.

First official results are yet to be reviewed and released to the market, together with the new tools for analysing this data and integrating it into the standard media planning and evaluation process.

5. Module 6 – Modelling audience accumulation for magazines

Introduction

The main objective for this module was to build an accumulation model for the magazines that would allow time planning based on issue readership estimates and its build up over the time, starting from the week of the release and growing up to the total estimated (accumulated) issue readership. It means that we needed to create an accumulation model which would allow us to estimate accumulating readership for an issue, week by week.

Cover recognition questions introduced in the new main NRS study were used as a basis for modelling the audience accumulation curves for magazines.

Our recommended approach for this task is a "multi-step" process, that builds on the final estimates created by the SIR model (process explained in the previous section; Module 5), which represent the maximum readership for a title, and then calculates readership "drop-offs" in periods leading up to the maximum readership. In order to calculate this, we have built an accumulation formula by fitting a "survival function" onto the accumulation pattern observed on the survey.

Method

The first step was to assess the available sample and make the decision on how the accumulation data collected on the survey will be used in the model.

Due to the "rolling" nature of the survey, and therefore the data collected for each period, we are able to collect reasonably large samples for each of the periods. Our original idea was to use different periods, across all issues, as individual samples, regardless of the title's periodicity. However, due to unstable publication schedules, this approach had to be scrapped, and we had to fit the accumulation curves to readership data split by week of recollection (i.e. read within one week of publication, read within two weeks of publication etc.).

The final dataset containing cover readership was firstly weighted at a total level. At the next step, the exact period between the publication and the time of appearance in the questionnaire was calculated for each surveyed issue. These periods were then recoded into full weeks since publication, creating an accumulated readership estimate for each particular "weeks passed" data point.

Based on this data, we then calculated accumulated readership across "weeks passed" points for each title, using four regular publication periods as the maximum point for each title (e.g. for monthlies, we used 17 weeks since publication as the maximum valid age for reading a title).

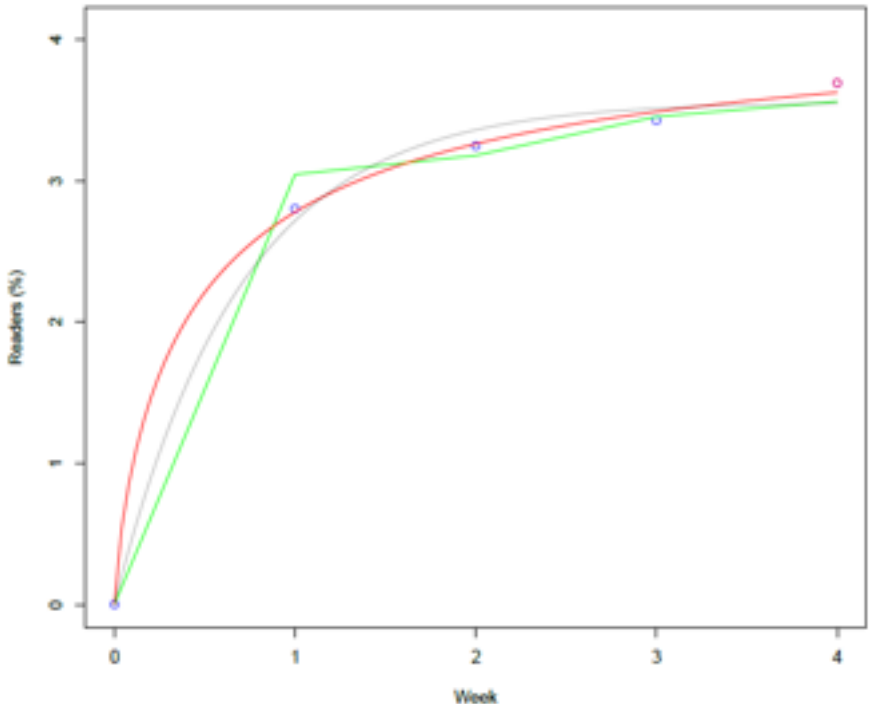
The points were then first evaluated at face value, to see if the accumulation patterns were sensible. As expected, titles that had reasonably large samples showed sensible accumulation patterns, with some memory decay present at 3rd and 4th oldest covers for monthlies, bimonthlies and quarterlies. All in all, data was deemed sufficient to proceed.

Our next step was to determine which point represented the maximum readership for each title. In cases where a consistent pattern of memory decay was present, we disregarded the data after the maximum reach point, where memory decay became larger than the audience build up. In our initial runs of the fitting function we saw that this presented a challenge, because a large number of titles recorded outlying readership levels in some of the weeks before a natural accumulation started being outweighed by the memory decay. To address this, we chose the maximum observed point based on the moving average of accumulated readership, where the moving average was constructed for each point by averaging the readership at that point, and the readership preceding and following that week. This proved effective, and allowed to build accumulation curves for most of the titles.

The final selection of the accumulation curves was performed by fitting a range of commonly used survival functions to the observations and examining which function was the best fit for each title. In the end, after analysis of all the fitted models, it was decided to proceed with the log-logistic function for most cases. For cases where the log-logistic function was proved to be unfeasible, we fitted a negative exponential function. In a small number of cases, a viable accumulation function could not be applied.

Finally, each survival function was rescaled to provide a factor of 1 when reaching the 4 regular publication periods for each title. This enabled us to directly apply the accumulation function to the SIR data to get the decrease of readership in the preceding periods.

Figure 10: Accumulation curve showing the moving average across weeks (green), the fitted log-logistic curve (red), and the alternative negative exponential curve (grey). The points represent observed readership across weeks



Conclusions

Most titles produced sensible accumulation curves, however, for some titles the sample sizes were insufficient, which as results introduced too much variance to be able to model a viable accumulation curve. In order to still produce a workable result for such titles we have applied the accumulation curve calculated on a group level, using title's corresponding genre and periodicity type.

As with the SIR modelling, some challenges presented themselves when analysing data after concluding the fieldwork. Firstly, although accumulation patterns were sensible, due to irregular publication schedules, samples for different "weeks passed" groups proved to be vastly different. The biggest problems remained with the issues that were "rushed" in publication, where special edition issues pushed regular issues out of the survey, reducing the amount of data available at certain position. In addition, due to sampling not being consistent across all weeks, some publication periods were left with significantly larger samples compared to other periods. This forced us to control for sample differences more rigorously, and, in addition, moved us from an approach of observing each title in its' own periodicity context, to a context of analysing data across weeks for each title. In the end, this approach proved fruitful, and could be implemented for all titles in the study.

Building on our previous experiences in creating similar models, accumulation modelling based on cover recognition data once again proved to be an effective way of obtaining accumulation patterns for magazines, regardless of size.

References

Cools B., Eeckhout l., Peeters S. & Raes Y. (PDRF 2013): MEASURING THE DIGITAL VERSIONS AND THE WEBSITE OF PRESS TITLES: THE BELGIAN EXPERIENCE

Frankel M.R., Baim J., Galin M.,. Agresti J., Mediamark Research Inc (WRRS 2007): ISSUE SPECIFIC ESTIMATION – MATHEMATICAL AND STATISTICAL ISSUES, PROCEDURES AND MODELS

Peeters S., Schrember J.S., Eeckhout L, Cools B (PDRF 2011): CAN COVER RECOGNITION REPLACE RECENT READING

Page K. (PDRF 2015): WHAT'S NEW? THE PDRF WORLDWIDE REVIEW OF AUDIENCE MEASUREMENT