# Opportunity To See Advertising (OTSA) in newspapers and magazines

Irena Petric (NOM), Alke Bassler (GfK Intomart) and Kenneth Gopal (ABF Research)

## 1. Introduction and Background

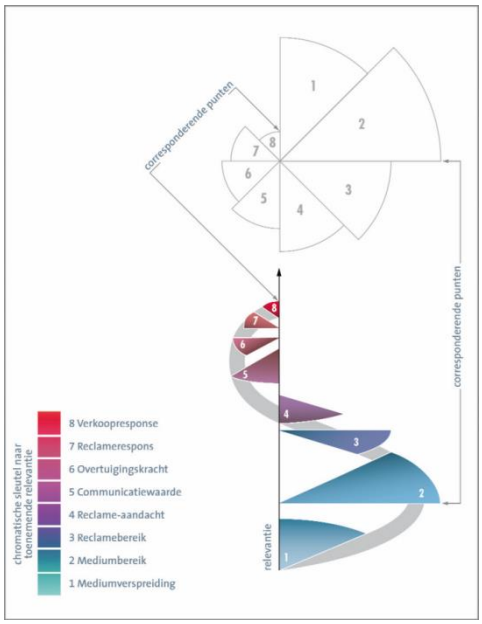
In 2010 the Dutch Advertisers Association BVA and Association of media-buying agencies PMA stated in a strategic paper that all the currency audience surveys in The Netherlands should publish metrics at the same level: the level of advertising reach. This would not only bring all the surveys to the same level of information in the ARF media model, but would also make the different currencies more comparable to each other. Newspaper and magazine readership surveys in most of the countries measure readership at the level of medium reach (Average Issue Readership). For the JIC that is responsible for the national readership survey, NOM, this meant that we had to think about how to obtain information about advertising reach of newspapers and magazines. We knew it was not possible to include questions on advertising reach into the currency survey as the length of the questionnaire does not allow for it. It is also questionable whether it is necessary to measure advertising reach on a continuous basis.

We looked for inspiration at the Norwegian newspaper survey (Futsaeter et al, WRRS 2009; Sandvik et al., PDRF 2011). In Norway they use a prediction model for calculating "extended currencies" for newspapers, a figure which goes beyond AIR. In their Media Effect Pyramid which is based on the ARF media model, they take a step up from Vehicle Exposure to Advertising exposure. In this way newspapers can compete beter with TV and radio data which report their currencies at this level.

Choosing modeling as a solution for taking a step up from Vehicle Exposure to Advertising Exposure for newspapers and magazines seems like an obvious choice. However we chose different independent variables to feed into the model than in the Norwegian example (see The basis for prediction models).

In the following paragraphs we will first explain how we obtained the basis for the prediction model for advertising exposure; then the obtained models will be described. Finally, we will show the results of the models and elaborate on the future steps.

Figure 1: ARF media model (the Dutch version translated by Kobalt media buying agency): from step 2 to step 3.

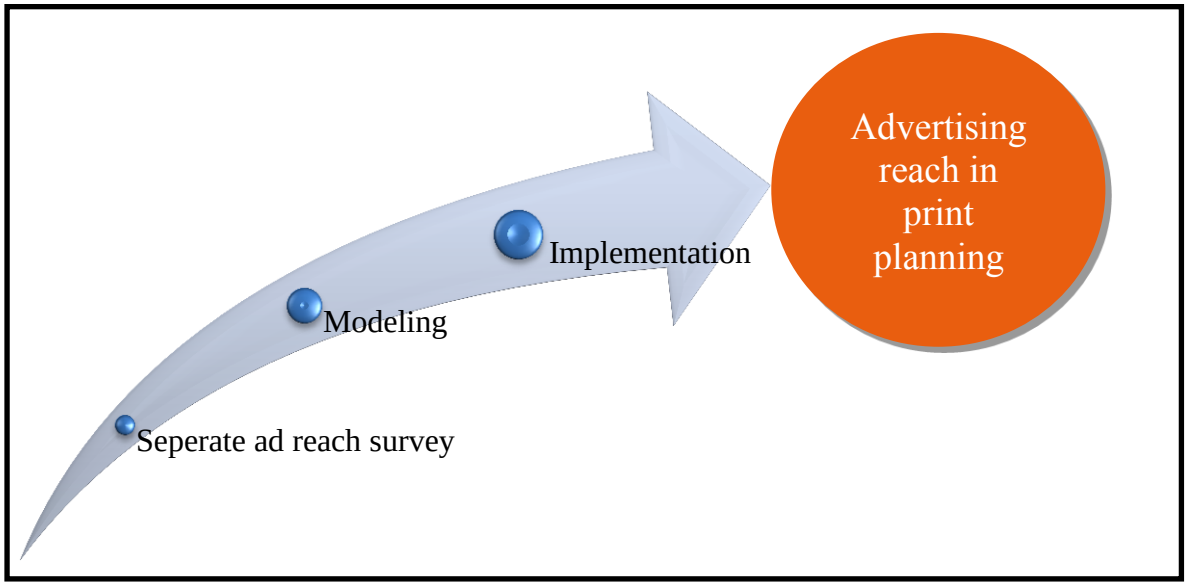


**2. The basis for the prediction models**

The following design was chosen. First, a survey, separate from the national readership survey, was set up in which the reach of 259 ad pages was measured. On the basis of the results prediction models were calculated. If the models would be technically approved, they could be implemented in the print planning software. The design of the project can be seen in Figure 2.

The biggest difference with the Norwegian newspaper survey is the selection of the independent variables that were considered for the model. In the Norwegian prediction model, only two independent variables were chosen: claimed reading of 8 newspaper editorial categories and amount of reading. In our OTSA survey, three types of independent variables were included: reader characteristics, title characteristics and advertisement characteristics. Also the relations between reader - title and reader- product were taken into account as independent variables. It was especially important to include also variables that are connected to the advertisement and product category, as they are known to influence advertising exposure. Advertising reach was defined as "having seen an advertising page" (advertising exposure).

Figure 2: Design of the Opportunity To See Advertising (OTSA) project.




**3. Survey design**

Advertising reach was measured for 18 titles (13 magazines, 5 newspapers), representative for the titles that are measured in NOM Print Monitor. Per title three issues were measured. Per issue 5 advertisements were selected to be measured; they were spread across different product categories and different types of advertisements (theme, introduction, action etc.). The position of the advertisement in the title was also taken into account (first section of a newspaper vs. other sections, left page / right page / cover etc).

The fieldwork was carried out in May through August 2012. From the GfK Online Access Panel respondents were recruited who had participated in the national readership survey. By using those respondents, it was possible to identify readers for the titles in the OTSA-survey and to use information on their reading behaviour that had been collected in the readership survey. In contrast to studies that only include frequent readers, this study was carried out among infrequent as well as frequent readers.

Respondents received an invitation by e-mail to fill in the online questionnaire; the questionnaire consisted of three parts:

1. Screen: reading the most recent issue

   Only respondents who had seen the most recent issue qualified for participation in the survey. To confront respondents with older issues would mean they could have forgotten that they saw the ad in a specific issue. Therefore, the surveys were started before the next issue was to be published (e.g. the surveys for daily newspapers were conducted the day after, surveys for monthlies one and a half week to a month after the publication).

2. Independent variables: questions about reading behaviour, attitude towards product category

   Because we recontacted respondents of the readership survey, many variables were already known, so the questionnaire could be kept short – 3 minutes on average.

3. Dependent variable: opportunity to see advertising

   Respondents were shown the spread which contained the ad page. They were asked whether they had opened the spread. If so, they were asked whether they had looked at the ad page we selected for the survey.

The 18 titles were selected from different sorts of publications: women's monthlies en weeklies, automotive magazines, tv-guides, sport's magazines, news weeklies, gossip, national and regional daily newspapers. From each title, three issues were selected. Within the issue, ad pages were chosen for the most common type of ads: for magazines whole advertising pages were selected, while for newspapers also smaller formats were measured. Ads for the following product categories were included: personal care/beauty, fashion, restaurants/tourism, automotive/transport, food, telecom/computers, retail, home & garden, consumer electronics, miscellaneous. Per issues four to five ad pages were selected, depending on how many advertisements there were that fit the selection criteria.

Each survey (18 titles x 3 issues = 54 surveys) was completed by 150 respondents, who answered the OTSA-questions for four to five ad pages. This resulted in 8,226 complete questionnaires and 38,166 cases. Because there is overlap in the titles that respondents read, respondents received surveys for more than one title (at different times). The response among the screened in respondents was 29%.

The cases in the survey of each title were weighted to fit the general profile of the title as measured in the national readership survey. The weighting correct for sex, age and reading frequency. Because these variables had been used to stratify the sample selection, the unweighted data already matched the target very well (Net Effective Base >95%). The only exceptions were the two regional newspapers, where the number of infrequent readers was underrepresented (Net Effective Base 64% and 76%).

## 4. The models

The data were used to build two models that predict the probability a reader has seen an ad page: one for daily newspapers and one for magazines. The models were built using logistic regression.

The dependent variable in the analyses was OTSA: ad page seen or ad page not seen. The independent variables consisted of characteristics of the reader, title, advertisement and their interrelations. Most of the independent variables are applicable to newspapers as well as magazines, some were only used in the analysis for either medium type. 19 independent variables were included in the analyses for newspapers and 21 for the analyses of magazines; when necessary, nominal and ordinal level variables were dummy coded. The candidate variables were used in a stepwise forward logistic regression. Table 1 shows the independent variables that were included in the analyses.

**Table 1. Independent variables used in logistic regression**

| Nr. | Cluster | Variable | Model | Note |
|---|---|---|---|---|
| 1 | Reader | Sex | newspapers/magazines | |
| 2 | | Age group | newspapers/magazines | 5 age groups |
| 3 | | Level of education | newspapers/magazines | |
| 4 | | Life cycle | newspapers/magazines | |
| 5 | | Position in household | newspapers/magazines | |
| 6 | | Affinity for advertising | newspapers/magazines | |
| 7 | Title-reader relationship | Number of pick ups | magazines | |
| 8 | | Moment of reading | magazines | weekend or work week |
| 9 | | Amount of reading | newspapers/magazines | |
| 10 | | Time spent reading | magazines | |
| 11 | | Reading location | magazines | |
| 12 | | Engagement | magazines | |
| 13 | | Reading frequency | newspapers/magazines | |
| 14 | | Source of copy | newspapers/magazines | |
| 15 | Advertisement | Product category | newspapers/magazines | |
| 16 | | Section number | newspapers | first part of first section vs. rest first section vs. not first section |
| 17 | | Format | newspapers | |
| 18 | | Above/below fold | newspapers | |
| 19 | | Colour | newspapers | |
| 20 | | Position | newspapers/magazines | left/right/inside cover/(inside) back page |
| 21 | | Adjancency editorial page | newspapers/magazines | next to editorial page or not; for newspapers also on editorial page |
| 22 | | Type of campaign | newspapers/magazines | Action, Introduction, Branding, Topical |
| 23 | | Sample/coupon | magazines | |
| 24 | Interest in product | Interest in product category | newspapers/magazines | |
| 25 | | Attention to advertising for product category | newspapers/magazines | |

For newspapers, 11 variables are necessary to predict the probability an ad page is seen. Table 2 lists the model variables and the effect each variable has on OTSA for newspapers. Variables are listed in order of the contribution to the model, i.e. the variable with the largest effect – amount of reading – is on position 1.

**Table 2. Variables in newspaper model**

| Nr. | Variable | Effect |
|---|---|---|
| 1 | amount of reading | more read => higher OTSA |
| 2 | on/next to editorial page | ad on editorial page => higher OTSA |
| 3 | attention to advertising for product category | more attention => higher OTSA |
| 4 | section number | first section => higher OTSA |
| 5 | colour | full colour => higher OTSA |
| 6 | position: left/right/back page | back page => highest OTSA |
| 7 | affinity for newspaper advertisement | higher affinity => higher OTSA |
| 8 | level of education | low to middle level => highest OTSA |
| 9 | product category | retail => highest OTSA |
| 10 | age group | older => higher OTSA |
| 11 | type of campaign | branding campaign => highest OTSA |

The modeling for magazines initially resulted in a model with 22 variables. 7 of these variables were based on questions not yet included in the national readership questionnaire, describing title-reader relationship:

- Time spent reading
- Engagement
- Location (dummy coded, 4 variables)
- Moment of reading (weekend vs. other days)

If these variables were essential to the model, new questions per title would have to be introduced to the readership questionnaire, which would mean the questionnaire would take longer to fill in. To investigate how essential these 'missing'

variables are, a second model for magazines was constructed, not using these variables in the logistic regression. The second modeling resulted in a model that was able to predict OTSA almost as well as the first model. It was decided that the incremental accuracy of the more detailed model was not worth adding questions to the readership survey.

The final model for magazines is shown in Table 3.

**Table 3. Variables in magazine model**

| Nr. | Variable | Effect |
|---|---|---|
| 1 | amount of reading | more read => higher OTSA |
| 2 | attention to advertising for product category | more attention => higher OTSA |
| 3 | source of copy | elsewhere => highest OTSA |
| 4 | affinity for magazine advertisement | higher affinity => higher OTSA |
| 5 | interest in product category | higher interest => higher OTSA |
| 6 | level of education | low to middle level => highest OTSA |
| 7 | age group | 50+ => highest OTSA |
| 8 | type of campaign: action | action => lower OTSA |
| 9 | position: left/right/back page | back page => highest OTSA |
| 10 | type of campaign: topical ad | topical ad => higher OTSA |
| 11 | product category | fashion => highest OTSA |
| 12 | next to editorial page/ad spread | next to editorial page => higher OTSA |
| 13 | sex | female => higher OTSA |
| 14 | main shopper | not the main shopper => higher OTSA |
| 15 | reading frequency | less than half => higher OTSA |

To evaluate the performance of the models, we looked at

- model accurancy
- cross validation
- fit of modeled vs. measured data

*Model accuracy*

The measurement of OTSA itself, without modeling, can provide the market with valuable information: how many readers see an ad, given its position in a magazine and other characteristics. Because of its value, this information has been collected by publishers since many years. In this particular study, an average ad page in a newspaper has been seen by 62% of the readers and an average ad page in a magazine by 57%. Media planners could use this information to adjust the average issue readership figures. However, on the level of the individual reader, they would be wrong in almost half of the cases: e.g. for newspapers, they would be right for 62% of the readers who had seen the ad page and for 38% of the readers who hadn't, resulting in a correct prediction of 53% of the readers. We compared this 'random accuracy' to the model accuracy. Model accuracy is the percentage of cases in which the modeled probability corresponds with the measured data. The measurement of 0 (ad page not seen) or 1 (ad page seen) was compared to the modeled OTSA score (below 50% = 0; 50% or above =1). The model accuracy for newspapers is 70%, an improvement of a factor 1,33 compared to the random accuracy. The model accuracy for magazines is 65%, an improvement of 1,28 compared to random accuracy.

*Cross validation*

Another test was the 75/25 cross validation, where three quarters of the sample were used to construct the model (training sample), which was then applied to the last quarter of the sample (validation sample). The model accuracy of both samples was compared. Cross validation is a test to see whether the model is valid for data which have not been used to construct the model. For newspapers as well as magazines, the improvement factor in the validation sample matched the improvement factor in the training sample.

*Fit of modeled vs. measured data*

We also compared the findings of the measurement to the results based on the model scores. This was done for different groups of publications and for different target groups. Figure 3 and Figure 4 show the comparisons for newspapers and for magazines for all respondents. When comparing the data for all titles – the 5 newspapers and the 13 magazines respectively – the figures match perfectly, as they should, because the measurement was the basis for the models. The question of course is, how well do the model scores – based on the measurement of all newspapers and on all magazines – fit a subset of those publications.

For national newspapers, the model predicts a slightly lower probability for seeing an ad page than the actual data, whereas there is a small overestimation of OTSA for regional newspapers. The difference between modeled OTSA and measured data is somewhat larger for magazines, especially for tv guides, in which the model predicts OTSA to be 2.7% higher than was actually measured. The larger difference is to be expected, as there is more variance between magazines in the percentage of readers that see an ad page. The modeled data follows the patterns for different publication groups well.

**Figure 3. Predicted vs. measured OTSA – national and regional newspapers**

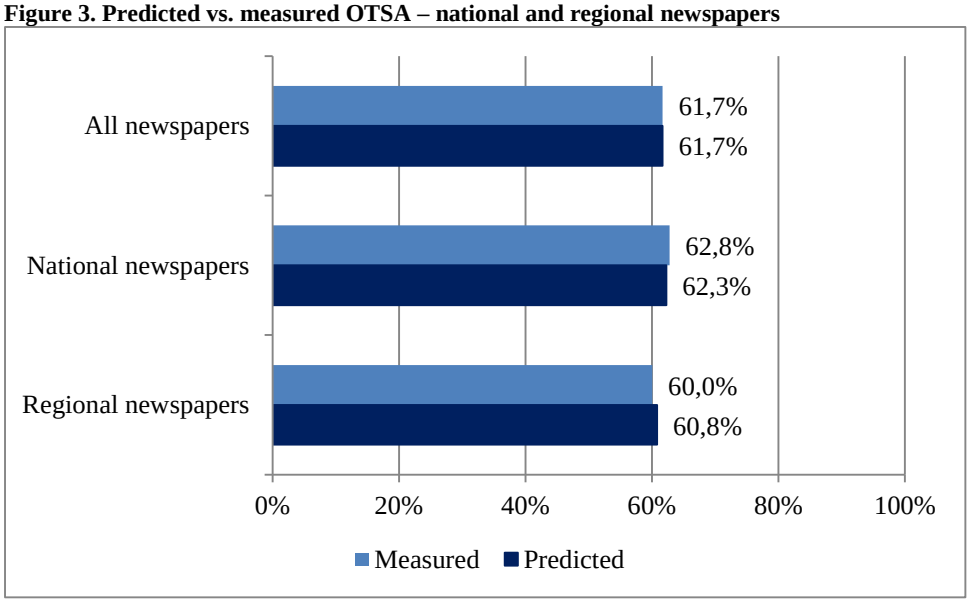
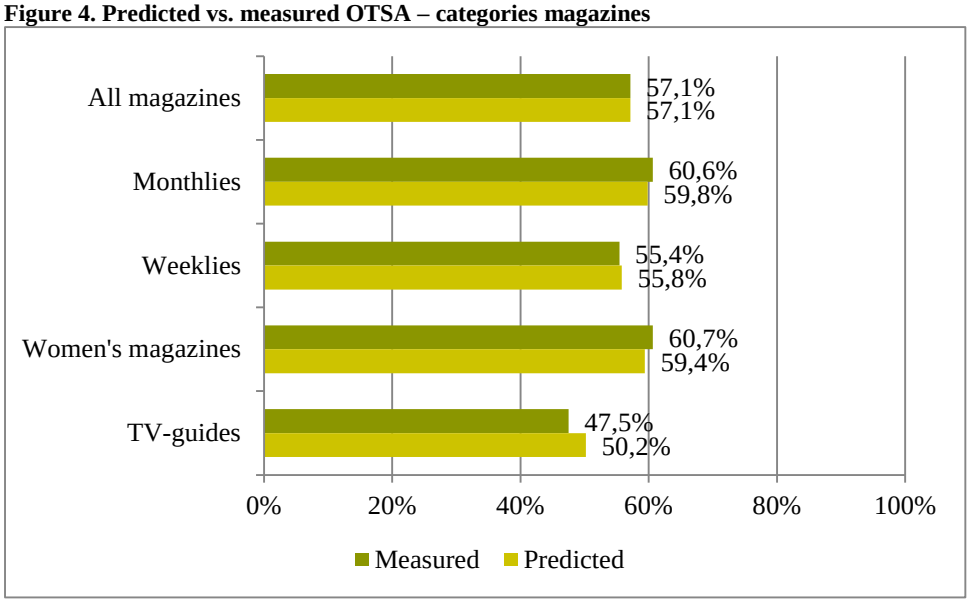
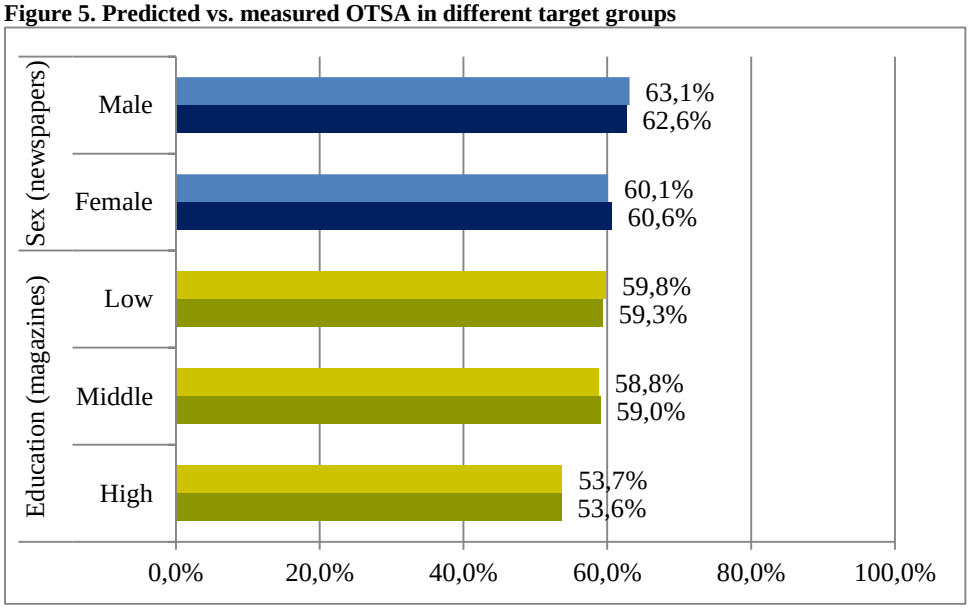



**Figure 4. Predicted vs. measured OTSA – categories magazines**




The measured and predicted results were also compared for different target groups. Again, the patterns found in the measurement were closely matched by the modeled data (see Figure 5). For newspapers, level of education and age are variables used in the model, the figures therefore show a perfect match. Sex is not a variable in the newspaper model, the model renders slightly lower ad page reach for men than was measured and therefore a slightly higher ad page reach for women. For magazines, sex and age are model variables, but level of education is not. The largest difference between modeled and measured data is found for readers with a low level of education.

**Figure 5. Predicted vs. measured OTSA in different target groups**




**6. Models applied to national readership survey**

The two prediction models were validated by the Technical Committee of NOM and found technically solid enough to be implemented. Next to the technical criteria we also took usability as a criterion to evaluate OTSA models (do the predicted values differ between the publication groups or product categories).

In a next step, we applied the models to all the titles in de national readership survey. In the separate ad reach survey, we found that there is differentiation in the results for different publications and target groups (shown in the figures in the paragraph above). We therefore expect to find differentiation in the national readership survey, when we apply the models to newspapers and magazines. Table 4 shows the predicted values for newspapers and magazines for different product categories, when applied to all 45 newspapers and about 140 magazines in the national readership survey. Especially for newspapers the range of ad page reach is big for different product categories.

On the basis of these results the Technical Committee has decided that the two OTSA models are both technically solid and suitable to be used in the media planning practice.

**Table 4. Average ad page reach per product category**

| Product category | Newspapers | Magazines |
|---|---|---|
| Beauty | 53% | 59% |
| Fashion | 59% | 61% |
| Tourism | 56% | 59% |
| Transport | 60% | 57% |
| Food | 48% | 60% |
| Telecom | 58% | 58% |
| Retail | 64% | 60% |
| Home and garden | 57% | 57% |
| Finance | 58% | 55% |
| Electronics | 56% | 58% |
| NEC | 52% | 58% |

**7. Next steps**

NOM has decided to include OTSA into its publication. At the moment we are talking to media planning software providers about how to implement the models into the print software. The next step is the implementation of the obtained OTSA prediction models in the market. This is certainly not the easiest part of the whole project; it is very important that all players in the market understand the implications of the new types of data for their everyday business.

Unlike the Norwegian newspaper industry, NOM decided that prediction of advertising exposure (OTSA) cannot be used as the new currency for newspapers and magazines. One of the most important characteristics of a currency is that it is based on representative, reliable and up-to-date survey results; OTSA is based on prediction models that are obtained from a relatively small single survey. We do not want to diminish the importance of obtaining these results, but we think that the results are not robust enough to be used as currency.

What OTSA prediction models certainly do is make print media planning more efficient and innovative. They also offer publishers more differentiation in their marketing activities. And finally, taking the step from Vehicle Exposure to Advertising Exposure for newspapers and magazines puts print into a better position in strategic comparison between different media types, especially with internet and television.

## References

Futsaeter, K.A., I. Sandvik & T. Ostnes: The New NRS in Norway – From AIR to Extended Currencies for Newspapers, WRRS 2009, I.

Sandvik, H. Holbaek – Hanssen, T. Ostnes, K.A. Futsaeter: Implementing Extended Currencies and Rethinking the Current Print Price Models, PDRF 2011