Estimating reach curves from one data point

Georg M. Goerg

Google Inc.

Last update: November 21, 2014

Abstract

Reach curves arise in advertising and media analysis as they relate the number of content impressions to the number of people who have seen it. This is especially important for measuring the effectiveness of an ad on TV or websites (Nielsen, 2009; PricewaterhouseCoopers, 2010). For a mathematical and data-driven analysis, it would be very useful to know the entire reach curve; advertisers, however, often only know its last data point, i.e., the total number of impressions and the total reach. In this work I present a new method to estimate the entire curve using only this last data point.

Furthermore, analytic derivations reveal a surprisingly simple, yet insightful relationship between marginal cost per reach, average cost per impression, and frequency. Thus, advertisers can estimate the cost of an additional reach point by just knowing their total number of impressions, reach, and cost.

A comparison of the proposed one-data point method to two competing regression models on TV reach curve data, shows that the proposed methodology performs only slightly poorer than regression fits to a collection of several points along the curve.

1 Introduction

Let $k+$ reach, r_k, be the percentage of the population that is exposed to a campaign at least k times. As usual, we measure impressions in gross rating points (GRPs), which is calculated as number of impressions divided by total (target) population multiplied by 100 (measured in percent).

Equipped with a functional form of the reach curve, a variety of quantities of interest can be computed, e.g., marginal cost per reach or maximum possible reach. Advertisers, however, often only have two points of the reach curve $r_k(g)$: $r_k(0) = 0$ and

$$r_k(G) = R \in [0, 100], \tag{1}$$

where $G \geq 0$ is the total GRPs and R is total reach. With this information alone one is tempted to use a linear approximation $r_k^{(1)}(g) = \frac{R}{G} g$. However, reach curves are not linear and in particular, the marginal reach per GRP would equal average reach per GRP (= 1/frequency); thus (1) alone is not helpful to get a better estimate of marginal GRP (and thus cost) per reach at $g = G$.

While the behavior of $r_k(g)$ around $g = G$ is in general unknown, the tangent at $g = 0$ can be approximated quite well: starting with no exposure, adding an infinitesimally small unit of GRPs (say ϵ) one reaches $\epsilon \cdot \iota$ % of the population, where $\iota = \iota(k)$ is the reciprocal of the expected number of impressions needed for the first person to see k impressions. One can lower bound ι by $1/k$. For $k = 1$, the bound is tight, $\iota = 1$; getting an exact expression of ι for $k > 1$ is ongoing research.[1] That is, for small g the reach curve can be approximated with a line through $(0, 0)$ with slope ι:

$$r_k(g) \approx g \cdot \iota \text{ for small } g. \tag{2}$$

Thus, approximately,

$$\lim_{G \to 0} \frac{\partial}{\partial g} r_k(g = G) = \iota. \tag{3}$$

Combining (1) with (3) allows us to estimate a two-parameter model.

Section 2 reviews parametric models for reach curves. Section 3 derives the parameter estimates based on

[1] In practice we found that $\iota = (k + \log_2 k)^{-1}$ gives good fits for several $k \geq 1$.

the total GRP and reach. Simulations and comparisons to full least squares estimates are presented in Section 4. Finally, Section 5 summarizes the main findings and discusses future work. Details on the TV reach curve data and analytical derivations can be found in the Appendix.

2 Reach curve models

Let $X \geq 0$ be the number of content impressions, e.g., TV shows, websites, or commercials. For a probabilistic view of reach curves, it is useful to decompose $k+$ reach as

$$\mathbb{P}(X \geq k, \text{reachable}) = \mathbb{P}(X \geq k \mid \text{reachable}) \cdot \mathbb{P}(\text{reachable}) \tag{4}$$

$$\Leftrightarrow r_k = p_k \cdot \rho, \tag{5}$$

where ρ is the maximum possible reach, and p_k is the probability of being reached k times, given that an individual is indeed reachable. This distinction allows us to model ρ and p_k with separate probabilistic models. Since reach is usually denoted in percent, we also use percent for maximum possible reach $\rho \in [0, 100]$, while we use proportions for $p_k \in [0, 1]$.

For further analytical derivations it is necessary to parametrize $p_k(g)$. Below we review two functional forms which are parsimonious (2 + 1 parameters), have excellent empirical fits, and lend themselves for simple analytical derivations.

2.1 Gamma-Mixture

Jin et al. (2012) propose a Poisson distribution for the impressions g, with an exponential prior distribution with rate β on the Poisson rate λ. This yields a model of the form

$$p_k(g) = 1 - \frac{\beta}{g + \beta}. \tag{6}$$

The exponential prior can be generalized to a $\Gamma(\alpha, \beta)$ distribution, which yields

$$r_k(g) = \rho \left(1 - \left(\frac{\beta}{\beta + g}\right)^{\alpha}\right). \tag{7}$$

By construction, (6) is nested in (7), which can be tested using a hypothesis test for $H_0 : \alpha = 1$.

2.1.1 Marginal reach

The derivative of (7) with respect to g equals[2]

$$\frac{\partial}{\partial g} p_k(g) = \frac{\alpha}{\beta} \left(\frac{\beta}{g + \beta}\right)^{\alpha + 1}, \tag{8}$$

with

$$\lim_{g \to 0} \frac{\partial}{\partial g} r_k(g) = \frac{\rho\alpha}{\beta}. \tag{9}$$

Eq. (9) has three degrees of freedom; since only two data points are available, one parameters has to be fixed. Given the nested structure of the exponential model, it is natural to set $\alpha \equiv 1$.

2.2 Conditional Logit

As an alternative we propose a logistic regression

$$\text{logit}(p_k(g)) = \beta_0 + \beta_1 \cdot \log g, \tag{10}$$

where $\text{logit}(p) = \log \frac{p}{1-p}$, and β_0 and β_1 are intercept and slope.[3] Using the logit inverse $\text{expit}(x) = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x}}$, Eq. (10) can be rewritten as

$$p_k = \text{expit}(\beta_0 + \beta_1 \log g) = \frac{e^{\beta_0 + \beta_1 \log g}}{1 + e^{\beta_0 + \beta_1 \log g}} \tag{11}$$

$$= 1 - \frac{1}{1 + e^{\beta_0} \cdot g^{\beta_1}} \tag{12}$$

$$= 1 - \frac{e^{-\beta_0}}{e^{-\beta_0} + g^{\beta_1}} \tag{13}$$

which shows similarity to (7). In fact, identifying $\beta \equiv e^{-\beta_0}$, both models coincide if $\alpha = 1$ and $\beta_1 = 1$, respectively. Again, this can be tested using a two-sided hypothesis test for $H_0 : \beta_1 = 1$.

The Logit conditional model can also be interpreted as the baseline Gamma mixture model with $\alpha \equiv 1$, but with transformed GRPs, $\tilde{g} = g^{\beta_1}$, in (7). Here β_1 can be interpreted as a parameter that measures the efficiency of GRPs: for $\beta_1 > 1$ GRPs are more efficient than baseline; for $\beta_1 = 1$ GRPs are spent according to the baseline model; and for $\beta_1 < 1$ are not spent as efficiently as expected. For an empirical estimates see Section 4.

[2] See Section B.1 for details.

[3] We deliberately do not use α and β to parametrize intercept and slope, as it is prone to confusion with the (reversed) roles of α and β in (8).

2.2.1 Marginal reach

The derivative of (11) with respect to g equals

$$\frac{\partial}{\partial g} p_k(g) = e^{\beta_0} \beta_1 \frac{g^{\beta_1 - 1}}{\left(e^{\beta_0} + g^{\beta_1}\right)^2}. \quad (14)$$

Here $\lim_{g \to 0} r'_k(g)$ falls into three cases:

$$\lim_{g \to 0} r_k(g) = \begin{cases} +\infty, & \text{if } \beta_1 < 1, \\ \frac{\rho}{e^{\beta_0}}, & \text{if } \beta_1 = 1, \\ 0, & \text{if } \beta_1 > 1. \end{cases} \quad (15)$$

Thus for the logit model one has to assume $\beta_1 = 1$ to use the linear approximation of $R(g)$ at $g = 0$ for 1+ reach.[4]

3 Methodology

Equipped with the two parameter model

$$r(g; \rho, \beta) = \rho \left(1 - \frac{\beta}{\beta + g}\right) = \rho \frac{g}{\beta + g} \in [0, \rho], \quad (16)$$

we can use the tangent approximation in (3) and total GRP and reach to estimate ρ and β. Note that $\beta \geq 0$ is a saturation parameter and controls how efficient GRPs are: for small β reach grows quickly with GRPs, for large β it grows slowly.

Its derivative equals

$$r'(g; \rho, \beta) = \rho \frac{\beta}{(\beta + g)^2}, \quad (17)$$

which at $g = 0$ evaluates to $r'(0) = \frac{\rho}{\beta}$.

This gives a system of two equations (maximum GRP and reach & marginal reach at 0) with two unknowns, $\rho \in [0, 100]$ and $\beta > 0$:

$$\frac{\rho}{\beta} = \iota \Leftrightarrow \rho = \beta \cdot \iota, \quad (18)$$

$$\rho \frac{G}{\beta + G} = R \Leftrightarrow \rho = \frac{R(G + \beta)}{G}. \quad (19)$$

First note that for 1+ reach, $\rho \equiv \beta$ since $\iota(k = 1) = 1$. Moreover, ρ in (19) satisfies $\rho \geq 0$ for all β, but it satisfies $\rho \leq 100$ only for $\beta \leq G \cdot \frac{100 - R}{R}$.

[4] For $k > 1$, the Logit model with $\beta_1 > 1$ might become useful as the marginal k+ reach for the very first impression is 0. However, one then has to estimate three parameters again, which is not possible without any further assumptions or more than one data point.

Solving for β and plugging in to $\rho = \rho(\beta)$ gives

$$\widehat{\rho} = \min\left(\frac{G \cdot R}{G - R/\iota}, 100\right), \quad (20)$$

$$\text{and } \widehat{\beta} = \begin{cases} \frac{\widehat{\rho}}{\iota} = \frac{G \cdot R/\iota}{G - R/\iota}, & \text{if } \rho < 100, \\ G \cdot \frac{100 - R}{R}, & \text{if } \rho = 100. \end{cases} \quad (21)$$

Condition $\widehat{\rho} \leq 100$ is equivalent to $G \leq \frac{100}{\iota} \frac{R}{100 - R}$; thus GRPs must be less or equal to a constant times the odds ratio of reach.

Plugging them back into (16) yields expressions for reach solely as a function of R and G (details see Appendix B). According to (21) we consider the two scenarios separately.

3.1 $\widehat{\rho} < 100$ case

Here

$$r(g) = \frac{G \cdot R \cdot g}{(G - g) \cdot R/\iota + g \cdot G} \quad (22)$$

with derivative

$$r'(g) = \frac{1}{\iota} \left(\frac{G \cdot R}{(G - g) \cdot R/\iota + g \cdot G}\right)^2. \quad (23)$$

At $g = G$ this evaluates to

$$r'(g = G) = \frac{1}{\iota} \left(\frac{R}{G}\right)^2. \quad (24)$$

Thus after G GRPs one additional GRP achieves an additional reach of (approximately) $\frac{1}{\iota}\left(\frac{R}{G}\right)^2$. Conversely, to get one additional reach point advertisers need approximately $\iota\left(\frac{G}{R}\right)^2$ additional GRPs. Since one GRP costs C/G, where C is total cost of the campaign, the marginal cost of one additional reach point is

$$\iota \left(\frac{G}{R}\right)^2 \times \frac{C}{G} = \iota \frac{C}{R} \frac{G}{R}. \quad (25)$$

Both (24) and (25) give two surprisingly simple, yet insightful identities which can be computed from total GRPs, reach, and cost: first, marginal reach per GRP equals a ι times squared average frequency $(= \frac{G}{R})$; secondly, marginal cost per reach equals

$$c'(r = R) = \iota \cdot \text{ cperp } \cdot \text{ frequency }, \quad (26)$$

where *cperp* is average *cost per effective reach point* (Rossiter and Danaher, 1998), and $c'(r)$ is the first derivative of cost as a function of reach, $c(r)$.

3.2 $\widehat{\rho} = 100$ case

If (20) leads to $\widehat{\rho} = 100$, then $\widehat{\beta}$ must be set to $G\frac{100-R}{R}$ to guarantee that $r(G) = R$. In this case (see Appendix B)

$$r(g) = \frac{g \cdot R}{G + (g - G) \cdot R/100} \tag{27}$$

with derivative

$$r'(g) = \frac{G \cdot (1 - R/100) \cdot R}{(G + (g - G) \cdot R/100)^2} \tag{28}$$

which at $g = G$ evaluates to

$$r'(g = G) = \frac{R}{G}\left(1 - \frac{R}{100}\right). \tag{29}$$

Again, this yields a simple identity for marginal reach as 1/frequency times the proportion of the population that has not been reached.

Thus marginal cost per reach is

$$c'(r) = \frac{G}{R}\frac{1}{1 - R/100} \times \frac{C}{G} = \text{cperp} \cdot \frac{1}{1 - R/100}. \tag{30}$$

Here, marginal cost per reach is average cost per effective reach point times a factor that is inverse proportional to the proportion of the population that has not yet been reached.

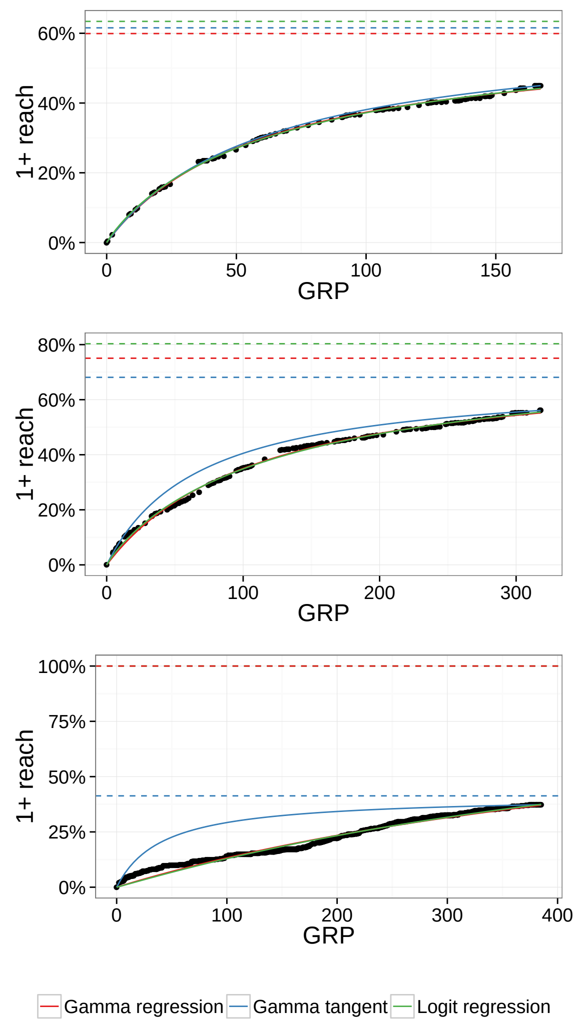


Figure 1: Sample reach curves with different degrees of Gamma tangent fit (bottom: lowest R^2; middle: median R^2; top: highest R^2).

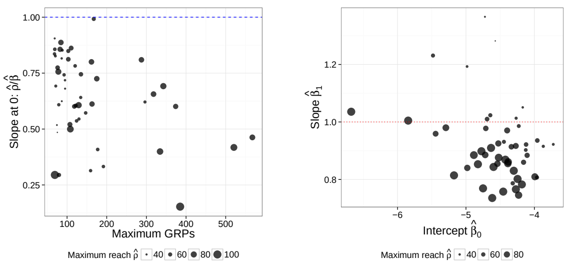
4 Applications

Here we compare the proposed one data point methodology to two competing regression methods using 1+ reach curves from a selection of 50 historical TV campaigns (See Appendix A for details on TV measurement and data processing). Note that for this comparison we do have several (typically hundreds of) data points along a single curve. Regression models use all the data points; the one data point methodology only uses the last data point. We evaluate the competing methods via typical model fitting metrics and ability to estimate marginal reach at $g = G$.

The two alternative regression methods are i) the Gamma mixture model, where ρ and β are estimated using non-linear least squares (set $\alpha \equiv 1$), and ii) the Logit model with logistic regression estimates for β_0 and β_1 (no restriction on β_1).

Figure 1 shows three reach curves with different degrees of fit: black dots are historical GRP and reach data, and colored lines are the model fits. The dashed horizontal lines represents the estimated $\widehat{\rho}$ for each method. The top shows a particularly good fit of all three models, the middle has a typical (median R^2) fit, and the bottom panel shows a reach curve where they do not coincide at all.

In fact, none of the proposed models provides a truthful representation of the GRP and reach relationship in the bottom panel. The raw data has several instances along the curve where reach flattens out already, but suddenly (at about 180 GRPs) it gains momentum and reaches again more people at a faster rate. One explanation could be that several creatives within the same campaign were shown sequentially; or the marketing strategy might have changed as a



(a) Gamma mixture model: estimated slope at zero $\widehat{\rho}/\widehat{\beta}$ (assuming $\alpha \equiv 1$). Blue, dashed line represents the Tangent model assumption of slope one.

(b) Logit model; red, dotted line represents $\beta_1 \equiv 1$.

Figure 2: Parameter estimates and inference about marginal reach at 0 GRPs.

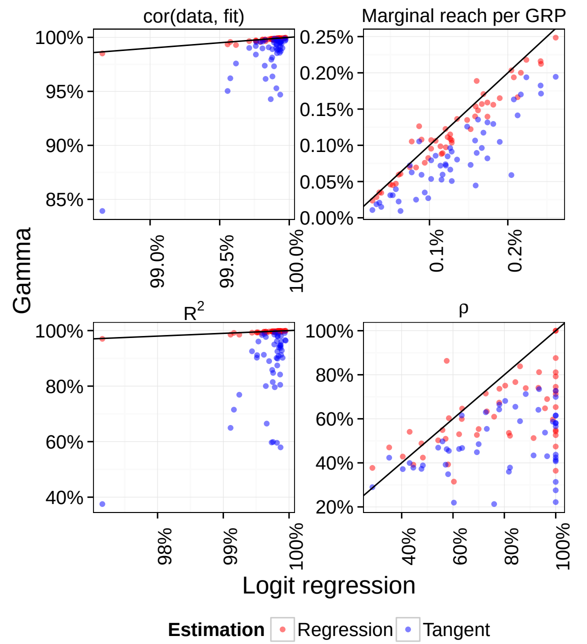


Figure 3: Comparing Gamma and Tangent estimates to – significantly better fitting – Logit model across several metrics.

consequence of the flattening out of the curve.

The bottom panel shows that in such cases the one data point methodology will fail since the campaign in fact consists of several sub-campaigns. However, while they give better fit, even the regression models are not really a good representation of the underlying GRP to reach dependency. For such campaigns a more general model which allows for multiple sub-campaigns should be used.

Figure 2 shows the estimated parameters for the Gamma mixture and the Logit model. Recall that in the Gamma Mixture model the slope at 0 equals $\alpha \cdot \frac{\rho}{\beta}$. Since $\alpha \equiv 1$ was fixed in the estimation, the estimated slope is simply the ratio $\frac{\widehat{\rho}}{\widehat{\beta}}$. Similarly to Fig. 1, the slope estiamtes in Figure 2a show that the one data point assumption (slope = 1) largely overestimates reach for small GRPs.

The Logit regression does not impose a $\beta_1 \equiv 1$ constraint, but all parameters were estimated from the data. Recall that β_1 can be interepreted as an efficiency parameter (see Section 2.2). According to $\widehat{\beta}_1$ in Fig. 2b about 80% of these campaigns do not use their GRPs as efficient as the baseline model would suggest.

Figure 3 compares the models according to several measures of fit. The Logit model stands out as a particularly good interpolation method (high R^2 and $cor(data, fit)$). Thus we use this – presumably closest to the truth – model as the baseline (x-axis) and check how the other two fare against it. Both the Gamma-mixture as well as the Tangent model infer much lower maximum possible reach. Note that the Logit regression estimate hits the boundary of $\widehat{\rho} = 100\%$ for 28% of the 50 campaigns.

The upper-right panel in Figure 3 shows that the Gamma and Logit regression marginal reach estimates coincide very closely, while the Tangent model[5] predicts lower marginal reach per GRP (below the 45° line), i.e., a flatter curve estimate. This is in agreement with the previous finding that the $\iota = 1$ slope at $g = 0$ is too optimistic; since the Tangent model always goes through the point (G, R) it must compensate the slope overestimation for small g, with a flatter curve for large G. As a consequence of the minimum restriction in (20), the tangent approximation yields some of the marginal reach per GRP estimates significantly below the 45° line.

Apart from these deviations, the scatterplots show that the proposed tangent method provides good estimates and useful inference.

[5] It is important to note that none of the reach curves have $\widehat{\rho} = 100$ in (20); the Tangent model estimates are thus all based on (24) with $\iota = 1$.

5 Discussion

In this work we show how to estimate the entire reach curve using only the total GRPs, reach, and cost. While a historical fit might not mimic the behavior for changes in future campaigns, it is very useful to estimate other quantities of interest from a historical campaign, such as maximum possible reach, marginal reach per GRP, or marginal cost per reach. Furthermore, we derive a simple, yet insightful equivalence between marginal cost per reach, average cost per GRP, and frequency.

Applications on a collection of historical TV reach curves show that the proposed method has good estimation properties and performs well against regression methods that use several data points at a time.

Acknowledgments

I would like to thank Christoph Best, Harry Case, Penny Chu, Tony Fagan, Jim Koehler, Raimundo Mirisola, Nicolas Remy, Mark Riseley, Sheethal Shobowale, and Xiaojing Wang for insightful discussions and constructive feedback. Special thanks go to Carsten Andreasen for getting the reach curve data.

References

Jin, Y., Shobowale, S., Koehler, J., and Case, H. (2012). The Incremental Reach and Cost Efficiency of Online Video Ads over TV Ads. Technical report, Google Inc.

Nielsen (2009). The Shifting Media Landscape - Integrated Measurement in a Multi-Screen World. Technical report, Nielsen.

PricewaterhouseCoopers (2010). Measuring the effectiveness of online advertising. https://www.pwc.com/en_GX/gx/entertainment-media/pdf/IAB_SRI_Online_Advertising_Effectiveness_v3.pdf.

Rossiter, J. and Danaher, P. (1998). *Advanced Media Planning*. Number v. 1 in Advanced Media Planning. Springer US.

TNS Gallup Denmark (2014). TNS TV/Radio Audience. http://tnsgallup.dk/markedsfokus/tv-radio-audience.

A Data Sources

For the model fit comparison in Section 4 we use TV measurement data from the Danish TV market. The raw data is based on a panel provided by TNS Gallup Denmark (TNS Gallup Denmark, 2014).

A.1 Panel recruitment

This panel consists of 1,000 households in Denmark, with approximately 2,250 panelists. Every household in this panel has a metering box and a remote control to log in when watching TV (including possibility to add guests).

Panelists have been recruited to be representative of Danish population, and weights are adjusted daily to calibrate panel for in and out-of-tab panelists. With a total population of about 5.6 million,[6] one panelist represents about 2,500 people.

A.2 Data selection and preparation

The metering box records TV viewing among panelists, and TV-stations report airing time of a campaign spots to TNS. GRPs per spot, and 1+-reach (in %) amongst others is calculated by TNS. To obtain the reach curves in Section 4 we compute cumulative GRPs and 1+-reach for each campaign. This data is then used to fit the presented models.

The data was collected on September 1, 2014 with a window of ± 2 months. The 50 campaigns we use here are based on a random subsample of the top quartile of all campaigns in the dataset. We use the top quartile to get campaigns with significantly large GRPs and reach.

[6] Source: http://denmark.dk/en/quick-facts/facts.

B Analytic derivations

B.1 Marginal reach

The derivative of (7) with respect to G equals

$$\frac{\partial}{\partial g} p_k(g) = \alpha \left(\frac{\beta}{g+\beta}\right)^{\alpha-1} \cdot \frac{\beta}{(g+\beta)^2} \tag{31}$$

$$= \frac{\alpha}{\beta}\left(\frac{\beta}{g+\beta}\right)^{\alpha+1} \tag{32}$$

$$= \frac{\alpha}{\beta+g} \cdot (1 - p_K(g)). \tag{33}$$

B.2 Plug-in β and ρ

B.2.1 $\rho < 100\%$ case

Plugging $\beta = \beta(G, R, k)$ and $\rho = \rho(G, R, k)$ back into (16) gives

$$r(g) = \rho \times \frac{g}{\beta + g} = \frac{G \cdot R}{G - R/\iota} \times \frac{g}{\frac{G \cdot R/\iota}{G - R/\iota} + g} \tag{34}$$

$$= \frac{G \cdot R}{G - R/\iota} \times \frac{\frac{g}{1}}{\frac{(G-g)\cdot R/\iota + g \cdot G}{G - R/\iota}} \tag{35}$$

$$= \frac{G \cdot R \cdot g}{(G-g) \cdot R/\iota + g \cdot G} \tag{36}$$

and derivative

$$r'(g) = \rho \frac{\beta}{(\beta+g)^2} = \iota \left(\frac{\beta}{\beta+g}\right)^2 \tag{37}$$

$$= \iota \left(\frac{\frac{G\cdot R/\iota}{G-R/\iota}}{\frac{G\cdot R/\iota}{G-R/\iota} + g}\right)^2 \tag{38}$$

$$= \iota \left(\frac{\frac{G\cdot R/\iota}{G-R/\iota}}{\frac{(G-g)\cdot R/\iota + g\cdot G}{G-R/\iota}}\right)^2 \tag{39}$$

$$= \frac{1}{\iota}\left(\frac{G \cdot R}{(G-g)\cdot R/\iota + g \cdot G}\right)^2 \tag{40}$$

Finally, the derivative at $g = G$ equals

$$r'(g = G) = \frac{1}{\iota}\left(\frac{G \cdot R}{(G-G)\cdot R/\iota + G \cdot G}\right)^2 \tag{41}$$

$$= \frac{1}{\iota}\left(\frac{R}{G}\right)^2. \tag{42}$$

B.2.2 $\rho = 100\%$ case

When $\rho = 1$ and $\beta = G\frac{100-R}{R}$ then

$$r(g) = 100 \frac{g}{\frac{G\cdot(100-R)}{R} + g} \tag{43}$$

$$= \frac{g \cdot R}{G + R/100 \cdot (g - G)}, \tag{44}$$

with derivative

When $\rho = 1$ and $\beta = G\frac{100-R}{R}$ the derivative simplifies to

$$r'(g) = 100 \times \frac{\beta}{(\beta+g)^2} \tag{45}$$

$$= 100 \times \frac{G\frac{100-R}{R}}{(G\frac{100-R}{R} + g)^2} \tag{46}$$

$$= 100 \times \frac{G(100-R)R}{(G(100-R) + g \cdot R)^2} \tag{47}$$

$$= 100 \times \frac{G(100-R)R}{(G \cdot 100 + R \cdot (g - G))^2} \tag{48}$$

$$= \frac{G \cdot (1 - R/100) \cdot R}{(G + (g-G)\cdot R/100)^2} \tag{49}$$