The Stata Journal (2017)
17, Number 2, pp. 358–371

Instantaneous geometric rates via generalized linear models

Andrea Discacciati
Karolinska Institutet
Stockholm, Sweden
andrea.discacciati@ki.se

Matteo Bottai
Karolinska Institutet
Stockholm, Sweden
matteo.bottai@ki.se

Abstract. The instantaneous geometric rate represents the instantaneous probability of an event of interest per unit of time. In this article, we propose a method to model the effect of covariates on the instantaneous geometric rate with two models: the proportional instantaneous geometric rate model and the proportional instantaneous geometric odds model. We show that these models can be fit within the generalized linear model framework by using two nonstandard link functions that we implement in the user-defined link programs `log_igr` and `logit_igr`. We illustrate how to fit these models and how to interpret the results with an example from a randomized clinical trial on survival in patients with metastatic renal carcinoma.

Keywords: st0478, log_igr, logit_igr, instantaneous geometric rate, generalized linear models, glm, survival analysis

1 Introduction

The geometric rate represents the average probability of an event of interest per unit of time over a specific time interval. Recently, Bottai (Forthcoming) showed that in the case of events that occur only once, such as death or first diagnosis of a disease, the geometric rate is a better measure of occurrence than the incidence rate. In the same article, Bottai proposed a regression method to model the conditional geometric rate given covariates. That method is based on applying quantile regression to a transformation of the time variable and is implemented in the user-written `grreg` command (Bottai 2015).

As the length of the time interval over which the geometric rate is defined shrinks to zero, we obtain the instantaneous geometric rate. This measure has a very intuitive interpretation because it represents the instantaneous probability of the event per unit of time.

In this article, we propose two models for the effect of covariates on the instantaneous geometric rate: the proportional instantaneous geometric rate model and the proportional instantaneous geometric odds model. We show that these models can be fit within the generalized linear model (GLM) framework (Nelder and Wedderburn 1972) by using two nonstandard link functions that can be easily programmed into the official Stata `glm` command (Guan and Gutierrez 2002).

The remainder of this article is organized as follows: In section 2, we briefly review how the instantaneous geometric rate is defined. In section 3, we show how to model the instantaneous geometric rate via GLM and present two user-defined link programs, `log_igr` and `logit_igr`. In section 4, we use data from a randomized clinical trial to illustrate some practical examples of how these link programs can be specified as an option of the `glm` command and how to interpret and present the analysis results. In section 5, we provide a summary.

2 Geometric rate and instantaneous geometric rate

In this section, we follow the description provided by Bottai (Forthcoming). Let T be a continuous random variable with support on $(0, +\infty)$ representing the time-to-event of individuals in some population, and let $S(t)$ be the associated survival function. The geometric rate over the time interval $(0, t)$ is defined as

$$g(0,t) = 1 - S(t)^{\frac{1}{t}}$$

and represents the average probability of the event per unit of time over $(0, t)$. The geometric rate between any two time points t_1 and t_2, such that $0 < t_1 < t_2 < +\infty$, is

$$g(t_1,t_2) = 1 - \left\{\frac{S(t_2)}{S(t_1)}\right\}^{\frac{1}{t_2-t_1}}$$

The limit of the geometric rate over shrinking time intervals $(t, t + \Delta t)$ gives the instantaneous geometric rate

$$\begin{aligned}
g(t) &\equiv \lim_{\Delta t\to 0^+} g(t, t+\Delta t) \\
&= \lim_{\Delta t\to 0^+} 1 - \left\{\frac{S(t+\Delta t)}{S(t)}\right\}^{\frac{1}{\Delta t}} \\
&= \lim_{\Delta t\to 0^+} 1 - \exp\left\{\frac{\log S(t+\Delta t) - \log S(t)}{\Delta t}\right\} \\
&= 1 - \exp\left\{\frac{\partial \log S(t)}{\partial t}\right\} \\
&= 1 - \exp\left\{-\frac{f(t)}{S(t)}\right\} \\
&= 1 - \exp\{-h(t)\} \qquad (1)
\end{aligned}$$

where $f(t)$ indicates the probability density function of T and $h(t) \equiv f(t)/S(t)$, the hazard function. The instantaneous geometric rate represents the instantaneous probability of the event per unit of time.

3 Instantaneous geometric rates via GLM

In this section, we show how instantaneous geometric rates can be estimated by GLM using nonstandard link functions. See Hardin and Hilbe (2012) for an exposition of GLM specifically targeted at Stata users.

Let t_i, $i = 1, \ldots, n$, be a sample of n possibly censored observations on the time variable, d_i be the event indicator variable (0 for a censored observation, 1 for an event), $\boldsymbol{x}_i = (x_{1,i} \ldots x_{p,i})'$ be a vector of covariates, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ be an unknown parameter vector.

3.1 Proportional instantaneous geometric rate model

We consider the proportional instantaneous geometric rates model

$$g_i(t|\boldsymbol{x}_i) = g_0(t) \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right) \tag{2}$$

By taking the logarithm of both sides of (2), we get

$$\log\left\{g_i(t|\boldsymbol{x}_i)\right\} = \log\left\{g_0(t)\right\} + \boldsymbol{x}_i'\boldsymbol{\beta}$$

and by taking the logarithm of (1), we get

$$\log\left[1 - \exp\left\{-h_i(t)\right\}|\boldsymbol{x}_i\right] = s(t;\boldsymbol{\gamma}) + \boldsymbol{x}_i'\boldsymbol{\beta} \tag{3}$$

where $s(t;\boldsymbol{\gamma})$ is a smooth parametric function of analysis time that depends on a vector of unknown parameters $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_r)'$.

To model the baseline log instantaneous geometric rate via $s(t;\boldsymbol{\gamma})$, we split each individual's follow-up into a number of intervals (or episodes) by choosing a fine grid of split points. After splitting the follow-up, let t_{ij} be the length of the jth time interval (the time at risk) relative to the ith individual, and let d_{ij} be the event indicator that takes value 1 if individual i develops the event in interval j, and 0 otherwise.

Following the same rationale behind parametric proportional hazard models (Royston and Lambert 2011, chaps. 4 and 7), (3) suggests using the following link function,

$$\eta_{ij} \equiv k(\mu_{ij}) = \log\left\{1 - \exp\left(-\frac{\mu_{ij}}{t_{ij}}\right)\right\} \tag{4}$$

where μ_{ij} is the expected value of d_{ij}, which is assumed to follow a distribution of the exponential family.

After suppressing the subscripts, the calculations to program the link function (4) are

$$\begin{aligned} \mu = k^{-1}(\eta) &= -t \log\{-\exp(\eta) + 1\} \\ \frac{\partial \mu}{\partial \eta} &= t \exp(\eta) \{-\exp(\eta) + 1\}^{-1} \\ \frac{\partial^2 \mu}{\partial \eta^2} &= t \exp(\eta) \{\exp(\eta) - 1\}^{-2} \end{aligned} \tag{5}$$

The following is the link program `log_igr`, contained in the `log_igr.ado` ado-file:

```
*! version 1.0.0 - 07dec2016
capture program drop log_igr
program define log_igr
        version 7
        args todo eta mu return

        if `todo´ == -1 { /* Title */
                global SGLM_lt "Log IGR"
                global SGLM_lf "log(1-exp(-u/$SGLM_p))"
                capture confirm numeric variable $SGLM_p
                        if _rc != 0 {
                                noi di as error "argument ($SGLM_p) to log_igr " /*
                                */ "link function must be a numeric variable"
                                exit 198
                        }
                exit
        }
        if `todo´ == 0 { /* eta = g(mu) */
                gen double `eta´ = log(-exp(-`mu´/$SGLM_p)+1)
                exit
        }
        if `todo´ == 1 { /* mu = g^-1(eta) */
                gen double `mu´ = -$SGLM_p*log(-exp(`eta´)+1)
                exit
        }
        if `todo´ == 2 { /* (d mu)/(d eta) */
                gen double `return´ = $SGLM_p*exp(`eta´)*(-exp(`eta´)+1)^(-1)
                exit
        }
        if `todo´ == 3 { /* (d^2 mu)/(d eta^2) */
                gen double `return´ = $SGLM_p*exp(`eta´)*(exp(`eta´)-1)^(-2)
                exit
        }
        noi di as err "Unknown call to glm link function"
        exit 198
end
```

To use this link, specify the `link(log_igr` *varname*`)` option in the `glm` command, where the existing numeric variable *varname* contains the time at risk, t_{ij}. See Guan and Gutierrez (2002) for a detailed explanation of how to program a custom link function.

3.2 Proportional instantaneous geometric odds model

We now consider the proportional instantaneous geometric odds model

$$\frac{g_i(t|\boldsymbol{x}_i)}{1-g_i(t|\boldsymbol{x}_i)} = \frac{g_0(t)}{1-g_0(t)}\exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right) \tag{6}$$

As we did in section 3.1, we write

$$\text{logit}[1-\exp\{-h_i(t)\}|\boldsymbol{x}_i] = s(t;\boldsymbol{\gamma}) + \boldsymbol{x}_i'\boldsymbol{\beta}$$

Therefore, the second proposed nonstandard link function is

$$\eta_{ij} \equiv k(\mu_{ij}) = \text{logit}\left\{1-\exp\left(-\frac{\mu_{ij}}{t_{ij}}\right)\right\}$$

and the necessary calculations to program it are

$$\begin{aligned}
\mu = k^{-1}(\eta) &= -t\log[\{\exp(\eta)+1\}^{-1}] \\
\frac{\partial\mu}{\partial\eta} &= t\exp(\eta)\{\exp(\eta)+1\}^{-1} \\
\frac{\partial^2\mu}{\partial\eta^2} &= t\exp(\eta)\{\exp(\eta)+1\}^{-2}
\end{aligned}$$

The following is the content of the `logit_igr.ado` ado-file, which contains the link program `logit_igr`:

```
*! version 1.0.0 - 07dec2016
capture program drop logit_igr
program define logit_igr
        version 7
        args todo eta mu return

        if `todo´ == -1 { /* Title */
                global SGLM_lt "Logit IGR"
                global SGLM_lf "logit(1-exp(-u/$SGLM_p))"
                confirm numeric variable $SGLM_p
                        if _rc != 0 {
                              noi di as error "argument ($SGLM_p) to logit_igr " /*
                              */ "link function must be a numeric variable"
                              exit 198
                        }
                exit
        }
        if `todo´ == 0 { /* eta = g(mu) */
                gen double `eta´ = logit(1-exp(-`mu´/$SGLM_p))
                exit
        }
        if `todo´ == 1 { /* mu = g^-1(eta) */
                gen double `mu´ = -$SGLM_p*log((exp(`eta´)+1)^(-1))
                exit
        }
```

```
        if `todo´ == 2 { /* (d mu)/(d eta) */
                gen double `return´ = $SGLM_p*exp(`eta´)*(exp(`eta´)+1)^(-1)
                exit
        }
        if `todo´ == 3 { /* (d^2 mu)/(d eta^2) */
                gen double `return´ = $SGLM_p*exp(`eta´)*(exp(`eta´)+1)^(-2)
                exit
        }
        noi di as err "Unknown call to glm link function"
        exit 198
end
```

Some notes are as follows:

1. Both models can easily accommodate time-varying covariates and time-dependent coefficients.
2. In (2), the exponentiated coefficients $\exp(\beta)$ are interpreted as instantaneous geometric rate ratios (IGRR), whereas in (6), they are interpreted as instantaneous geometric odds ratios (IGOR).
3. If the instantaneous geometric rates are proportional across different populations, the instantaneous geometric odds are not, and vice versa.
4. The inverse link function (5) is defined only for $\eta < 0$. This has two practical consequences. First, the default initial values $(\boldsymbol{\gamma}_0, \boldsymbol{\beta}_0) = (0, 0, \ldots, 0)$ used for the maximization of the log likelihood (Gould, Pitblado, and Poi 2010) are not feasible, because the log likelihood cannot be evaluated in $(\boldsymbol{\gamma}_0, \boldsymbol{\beta}_0)$. This can be solved by passing feasible initial values to `glm` or by specifying the `search` option (see [R] **maximize**). Second, the parameter space for $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ is bounded, which means the log likelihood is defined only within that parameter space. This introduces challenges in maximizing the log likelihood and may lead to failed convergence of the optimization algorithms, similarly to what happens to binomial models with a log link (Williamson, Eliasziw, and Fick 2013).

4 Example: Survival in metastatic renal carcinoma

We illustrate the use of the two proposed regression models using data from a clinical trial on 347 patients diagnosed with metastatic renal carcinoma (Medical Research Council Renal Cancer Collaborators 1999). The patients were randomly assigned to either interferon-α (IFN) or oral medroxyprogesterone (MPA). A total of 322 patients died during follow-up.

4.1 Data preparation

The numeric variable `survtime` represents the time in days to death or censoring, the binary variable `cens` indicates the death status (`0` = censored, `1` = death), and the variable `pid` contains the unique patient identifier.

First, we declare the data to be survival-time data with the `stset` command, and we rescale the analysis time from days to years with the `scale(365.24)` option.

Next, we split each patient's follow-up in intervals of length equal to one week using the `stsplit` command with the `every(‘=1/52’)` option and generate a new variable containing the time at risk within each interval (`risktime`).

Last, to model the baseline instantaneous geometric rate, we generate restricted cubic spline (RCS) transformations of analysis time, using the user-written `rcsgen` command (Lambert 2008). We use four knots, which by default are located at the minimum, maximum, and the 33rd and 66th centiles of the uncensored survival times' distribution. To do so, we add the `df(3)` and `if2(_d == 1)` options.

```
. use http://www.imm.ki.se/biostatistics/data/kidney
(Metastatic renal carcinoma trial. MRCRCC. Lancet. 1999, 353:14-7)
. stset survtime, failure(cens) id(pid) scale(365.24)

                id:  pid
     failure event:  cens != 0 & cens < .
obs. time interval:  (survtime[_n-1], survtime]
 exit on or before:  failure
    t for analysis:  time/365.24
--------------------------------------------------------------------------------
        347  total observations
          0  exclusions
--------------------------------------------------------------------------------
        347  observations remaining, representing
        347  subjects
        322  failures in single-failure-per-subject data
    375.687  total analysis time at risk and under observation
                                        at risk from t =         0
                              earliest observed entry t =         0
                                   last observed exit t =  6.209616

. stsplit click, every(`=1/52´)
(19,360 observations (episodes) created)
. generate risktime = _t - _t0
. rcsgen _t, df(3) if2(_d == 1) gen(_rcs)
Variables _rcs1 to _rcs3 were created
```

4.2 Proportional instantaneous geometric rates model

We fit a proportional instantaneous geometric rates model with the `glm` command with the `log_igr` custom link program. The `risktime` variable, which contains t_{ij}, is passed as an argument to `log_igr`.

We start by including the binary treatment indicator (`trt`) and the RCS transformations of analysis time (`_rcs1`, `_rcs2`, and `_rcs3`) in the model. The outcome variable `_d` contains the event indicator d_{ij}.

```
. glm _d i.trt c._rcs?, family(poisson) link(log_igr risktime) vce(robust) nolog
> search eform
initial:       log pseudolikelihood =     -<inf>  (could not be evaluated)
feasible:      log pseudolikelihood = -4804.4455
rescale:       log pseudolikelihood = -1959.6083
Generalized linear models                         No. of obs      =     19,707
Optimization     : ML                             Residual df     =     19,702
                                                  Scale parameter =          1
Deviance         =    3239.4169                   (1/df) Deviance =   .1644207
Pearson          =  124086.9279                   (1/df) Pearson  =   6.298189
Variance function: V(u) = u                       [Poisson]
Link function    : g(u) = log(1-exp(-u/risktime)) [Log IGR]
                                                  AIC             =   .1975652
Log pseudolikelihood =  -1941.70845               BIC             =  -191588.3
```

_d	exp(b)	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
trt						
IFN	.8371623	.0568225	-2.62	0.009	.7328824	.9562799
_rcs1	.9604894	.2909327	-0.13	0.894	.5304729	1.739089
_rcs2	1.308916	.8565102	0.41	0.681	.3630067	4.719643
_rcs3	.9010516	.2378547	-0.39	0.693	.5370987	1.511629
_cons	.7243848	.0642749	-3.63	0.000	.6087542	.8619789

The estimated IGRR comparing the two treatment groups (IFN versus MPA) is 0.84 (95% confidence interval: $[0.73, 0.96]$), constant throughout the entire follow-up. Under this model, the instantaneous yearly probability of death in the IFN group was estimated to be 16% lower than in the MPA group. We can predict the log instantaneous geometric death rate for the two treatment groups with the `predict` postestimation command.

```
. predict log_igr, xb
. generate igr = exp(log_igr)
```

In figure 1, we see that the instantaneous yearly risk of dying in patients on MPA decreased from about 75% to 25% over the 6 years of follow-up. Figure 1 also clearly exhibits the assumption of proportional instantaneous geometric rates in that the vertical distance between the two lines (on the log scale) is constant throughout the follow-up.

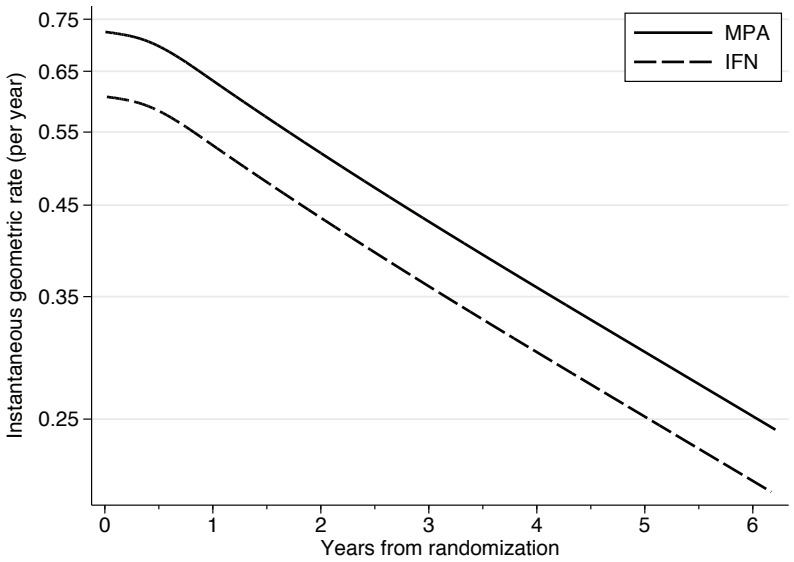

Figure 1. Predicted instantaneous geometric death rates for the two treatment groups from an instantaneous geometric proportional rates model. The vertical axis is on a log scale.

We now relax the assumption of constant IGRR. To do so, we add interactions (product terms) between `trt` and the three RCS transformations of analysis time.

```
. glm _d i.trt##c._rcs?, family(poisson) link(log_igr risktime) vce(robust)
> nolog search
initial:       log pseudolikelihood =     -<inf>  (could not be evaluated)
feasible:      log pseudolikelihood = -4804.4455
rescale:       log pseudolikelihood = -1959.6083
Generalized linear models                         No. of obs      =     19,707
Optimization     : ML                             Residual df     =     19,699
                                                  Scale parameter =          1
Deviance         =  3237.985686                   (1/df) Deviance =   .1643731
Pearson          =  122535.5358                   (1/df) Pearson  =   6.220394
Variance function: V(u) = u                       [Poisson]
Link function    : g(u) = log(1-exp(-u/risktime)) [Log IGR]
                                                  AIC             =    .197797
Log pseudolikelihood = -1940.992843               BIC             =  -191560.1
```

_d	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
trt						
IFN	-.3683698	.1914883	-1.92	0.054	-.7436799	.0069403
_rcs1	-.3103316	.3603476	-0.86	0.389	-1.0166	.3959368
_rcs2	-.2934956	.8245528	-0.36	0.722	-1.909589	1.322598
_rcs3	.1167311	.3344125	0.35	0.727	-.5387053	.7721675
trt#c._rcs1						
IFN	.7833079	.6631432	1.18	0.238	-.5164289	2.083045
trt#c._rcs2						
IFN	1.572779	1.383879	1.14	0.256	-1.139574	4.285131
trt#c._rcs3						
IFN	-.6156042	.5543412	-1.11	0.267	-1.702093	.4708846
_cons	-.2639627	.0874417	-3.02	0.003	-.4353452	-.0925802

```
. predict log_igr, xb
. generate igr = exp(log_igr)
. predictnl log_igrr = _b[1.trt] + _b[1.trt#c._rcs1]*_rcs1 +
> _b[1.trt#c._rcs2]*_rcs2 +  _b[1.trt#c._rcs3]*_rcs3
. generate igrr = exp(log_igrr)
```

The log-time dependent IGRR is obtained with the `predictnl` postestimation command and plotted in figure 2 after exponentiation, together with the instantaneous geometric death rates for the two treatment groups.

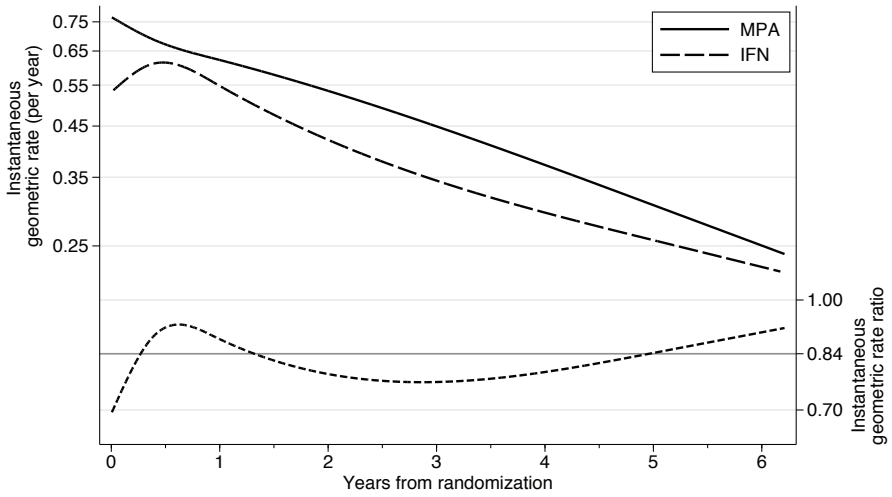

Figure 2. Predicted instantaneous geometric death rates for patients on MPA (solid black line) and IFN (long-dashed black line) and predicted time-dependent IGRR (short-dashed black line) (IFN versus MPA). The gray solid line indicates the time-fixed IGRR, equal to 0.84. The vertical axes are on a log scale.

When we inspect figure 2, it seems the assumption of constant IGRR throughout the follow-up is tenable. We can formally test this assumption by testing the coefficients of the interaction terms to be jointly equal to zero. This can be done with the `testparm` postestimation command.

```
. testparm 1.trt#c._rcs?
 ( 1)  [_d]1.trt#c._rcs1 = 0
 ( 2)  [_d]1.trt#c._rcs2 = 0
 ( 3)  [_d]1.trt#c._rcs3 = 0
           chi2(  3) =    1.43
         Prob > chi2 =    0.6983
```

From this output, we fail to reject the null hypothesis of proportionality of the instantaneous geometric rates (p-value = 0.6983).

4.3 Proportional instantaneous geometric odds model

To illustrate the proportional instantaneous geometric odds model, we now explore whether white cell count (`wcc`), a continuous prognostic factor, affects the treatment effect as measured by the IGOR. This analysis builds upon the findings reported by Royston, Sauerbrei, and Ritchie (2004), where they observed a beneficial effect of IFN—in terms of relative hazard—only among patients with a white cell count lower than about $10 \times 10^9\ L^{-1}$.

We include the treatment indicator, white cell count, their interaction term, and the three RCS transformations of analysis time as covariates. We specify the option `link(logit_igr risktime)` to fit a proportional instantaneous geometric odds model.

```
. glm _d i.trt##c.wcc _rcs?, family(poisson) link(logit_igr risktime)
> vce(robust) nolog
Generalized linear models                         No. of obs      =     19,707
Optimization     : ML                             Residual df     =     19,700
                                                  Scale parameter =          1
Deviance         =  3210.596989                   (1/df) Deviance =   .1629745
Pearson          =   119282.802                   (1/df) Pearson  =   6.054965
Variance function: V(u) = u                       [Poisson]
Link function    : g(u) = logit(1-exp(-u/risktime))[Logit IGR]
                                                  AIC             =   .1963057
Log pseudolikelihood = -1927.298494               BIC             =  -191597.4
```

_d	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
trt						
IFN	-1.674116	.5957372	-2.81	0.005	-2.841739	-.5064921
wcc	.0824596	.0453305	1.82	0.069	-.0063865	.1713058
trt#c.wcc						
IFN	.1620935	.0705864	2.30	0.022	.0237467	.3004403
_rcs1	.7416164	.8740937	0.85	0.396	-.9715757	2.454809
_rcs2	2.101511	1.771603	1.19	0.236	-1.370766	5.573789
_rcs3	-.8266814	.7022011	-1.18	0.239	-2.20297	.5496075
_cons	-.0688033	.4756726	-0.14	0.885	-1.001105	.8634979

Based on the p-value for the interaction term, we reject the null hypothesis of constant treatment effect throughout the observed range of white cell count (p-value = 0.022).

```
. predictnl log_igor = _b[1.trt] + _b[1.trt#c.wcc]*wcc, se(log_igor_se)
. generate igor = exp(log_igor)
. generate igor_lo = exp(log_igor - 1.96*log_igor_se)
. generate igor_hi = exp(log_igor + 1.96*log_igor_se)
```

The log IGOR comparing mortality among patients on IFN and patients on MPA as a function of white cell count can be obtained with the `predictnl` postestimation command and then plotted (figure 3).

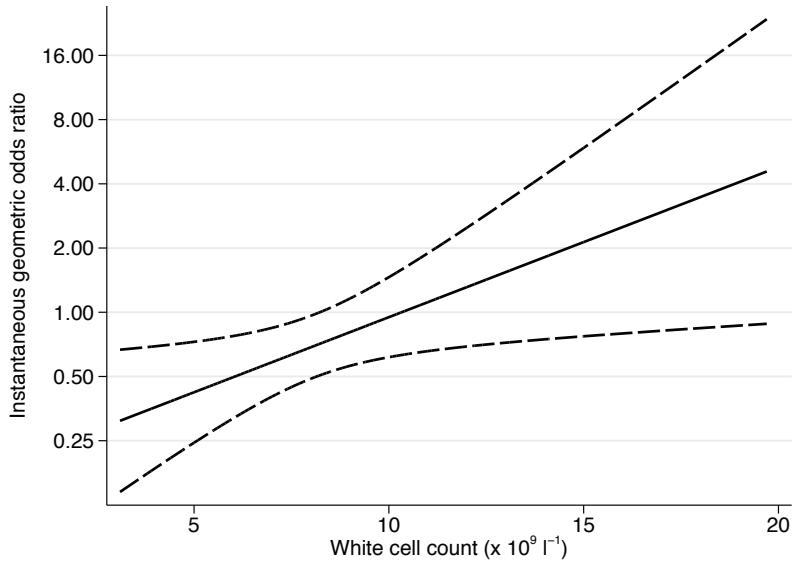

Figure 3. Predicted IGOR for IFN versus MPA (solid line) with 95% confidence interval (long-dashed lines) as a function of white cell count. The vertical axis is on a log scale.

The treatment effect seems to be largest among patients with a low white cell count. For example, the estimated IGOR for white cell counts of 4.9 and 13.7×10^9 L^{-1} (5th and 95th centiles of `wcc` distribution) were 0.40 (95% confidence interval: $[0.23, 0.72]$) and 1.72 (95% confidence interval: $[0.74, 4.04]$), respectively.

5 Summary

In this article, we proposed a method to model the effects of covariates on the instantaneous geometric rate within the GLM framework by using two nonstandard link functions. We showed how these link functions could be easily programmed into the `glm` command by creating two short, independent ado-files, `log_igr.ado` and `logit_igr.ado`.

Using data from a randomized clinical trial on survival in patients with metastatic renal carcinoma, we illustrated how to use these link programs and how to interpret results from the proportional instantaneous geometric rate model and the proportional instantaneous geometric odds model. We also demonstrated that a clear advantage of using `glm` to fit these models is that postestimation commands for `glm` are readily available.

In conclusion, the intuitive interpretation of the instantaneous geometric rate and the ease with which the proposed regression models can be fit in Stata make them a useful addition to the existing tools for the analysis of survival data.

6 References

Bottai, M. 2015. The grreg command for geometric rate regression. 2015 Nordic and Baltic Stata Users Group meeting proceedings. http://www.stata.com/meeting/nordic-and-baltic15/abstracts/materials/sweden15_bottai.pdf.

———. Forthcoming. A regression method for modelling geometric rates. *Statistical Methods in Medical Research*.

Gould, W., J. Pitblado, and B. Poi. 2010. *Maximum Likelihood Estimation with Stata*. 4th ed. College Station, TX: Stata Press.

Guan, W., and R. G. Gutierrez. 2002. Programmable GLM: Two user-defined links. *Stata Journal* 2: 378–390.

Hardin, J. W., and J. M. Hilbe. 2012. *Generalized Linear Models and Extensions*. 3rd ed. College Station, TX: Stata Press.

Lambert, P. 2008. rcsgen: Stata module to generate restricted cubic splines and their derivatives. Statistical Software Components S456986, Department of Economics, Boston College. https://ideas.repec.org/c/boc/bocode/s456986.html.

Medical Research Council Renal Cancer Collaborators. 1999. Interferon-α and survival in metastatic renal carcinoma: Early results of a randomised controlled trial. *Lancet* 353: 14–17.

Nelder, J. A., and R. W. M. Wedderburn. 1972. Generalized linear models. *Journal of the Royal Statistical Society, Series A* 135: 370–384.

Royston, P., and P. C. Lambert. 2011. *Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model*. College Station, TX: Stata Press.

Royston, P., W. Sauerbrei, and A. Ritchie. 2004. Is treatment with interferon-alpha effective in all patients with metastatic renal carcinoma? A new approach to the investigation of interactions. *British Journal of Cancer* 90: 794–799.

Williamson, T., M. Eliasziw, and G. H. Fick. 2013. Log-binomial models: Exploring failed convergence. *Emerging Themes in Epidemiology* 10: 14.

About the authors

Andrea Discacciati is a biostatistician in the Unit of Biostatistics at the Institute of Environmental Medicine at Karolinska Institutet in Stockholm, Sweden.

Matteo Bottai is a professor of biostatistics in the Unit of Biostatistics at the Institute of Environmental Medicine at Karolinska Institutet in Stockholm, Sweden.