RGR MS

Chris H. Wilson

March 2, 2019

### Key Points

1. Terminology: “linear difference” -> “linear growth”; “log measure” -> “log RGR”; “linear RGR != linear growth”
2. Be clear about the dimensional comparisons. The RGR measures are not dimensionally the same as linear growth.

2b) RGR measures and all comparisons really, sensitive to the delta.

1. In the absence of a time-varying model, estimating RGR at one or two points in time is practically useless, without super strong priors. Where we are comparing across species, this gets complicated. The problem is when we do NOT know WHERE on the growth curve our observations are coming from.
2. The conclusion is a bit more annoying. Linear growth is just as good or better for much of the curve, but this offers little comfort when we are making cross-species’ comparisons. If the goal is to estimate a treatment effect, linear growth at least is interpretable.

### Background and Rationale

Analyzing growth of individuals is fundamental in many areas of ecology and biology. A common situation is the need to compare multiple individuals across genotypes or species in experimental or observational settings where variations in initial sizes and environmental factors both contribute to observed variations in growth. In this setting, a common default practice is to re-express growth as a relative measure, dividing the growth increment by the initial size. In the limit as the time period goes to zero, this can be represented as

Without explicit specification of a time-varying dynamic, e.g. some kind of non-linear growth function, this representation corresponds to exponential growth. That is, the quantity obtained by integration of

over some time period, and given some initial size

is simply the familiar exponential equation

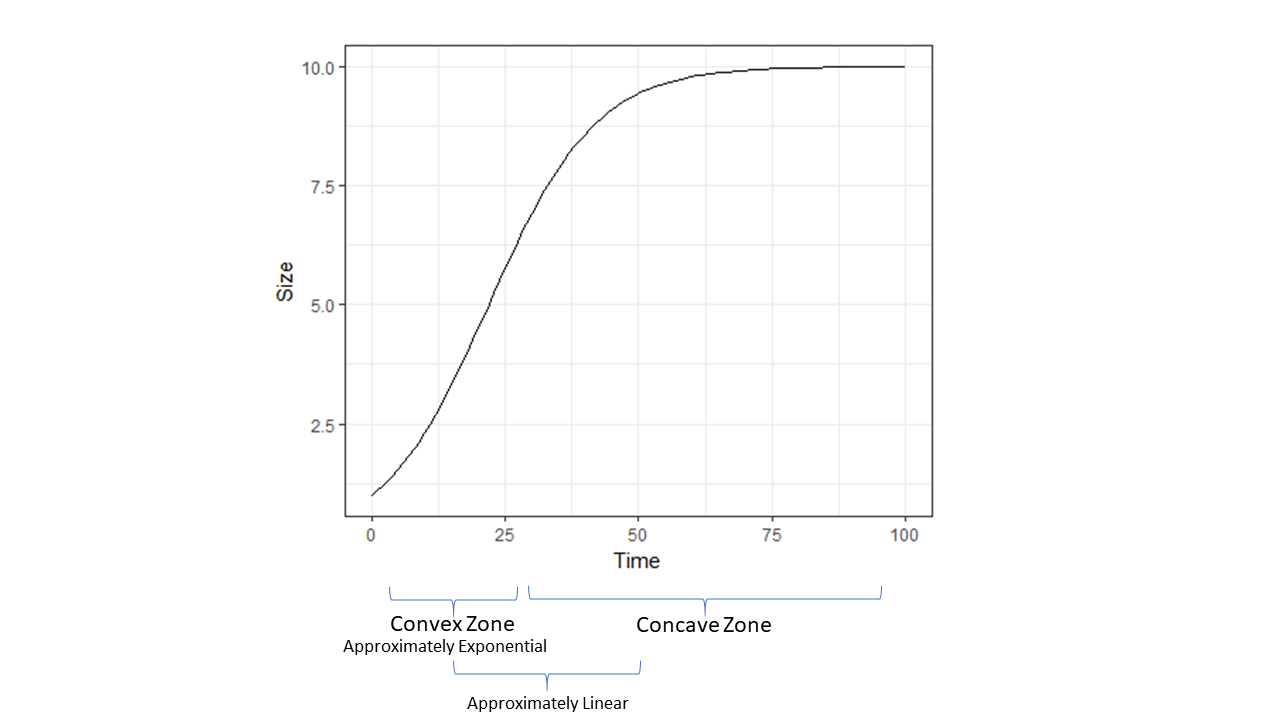
The quantity in equation 1 is often referred to as relative growth rate (RGR), and the usual method of quantification, hereafter the “log RGR” corresponds to the solution in 2, as is readily checked. The log RGR is, simply The log RGR is very frequently utilized as a default in place of taking the difference , hereafter “linear growth rate”, since it is seen as more effectively accounting for variations in initial size, given the understanding that size itself is a fundamental driver of subsequent growth. Note that the linear growth rate could also be divided by to return a “linear RGR”.

XXX et al. (2012) summarized several flaws of the log RGR and recommended instead to fit non-linear growth functions. The non-linear growth functions can then be differentiated with respect to size in order to obtain a superior RGR measure. I wholeheartedly concur with this advice. However, ecologists are often confronted with datasets where only 2 or 3 time periods are available, thus precluding effective fitting of non-linear functions. In this context, the log RGR is widely recommended.

In this note, I demonstrate that the linear measures (linear growth rate, and linear RGR) are in many cases superior to the log RGR. To be sure, these quantities should be seen as answers to subtley different questions. My intent is to highlight the underlying assumptions, and challenge the status of log RGR as a default in the data limited setting.

### Conceptual Overview

First, we assume a basic theoretical framework for growth: the sigmoidal curve. Nearly every biologically motivated growth model follows sigmoidal behavior. For instance, West et al. (2001) famously derived a sigmoidal equation for growth from metabolic scaling theory. Mechanistic models of photosynthesis and leaf area also result in sigmoidal growth (CITE). Although particular sigmoidal models can be challenged, the qualitative pattern is universal. Therefore, we will compare log RGR and linear growth rate to a sigmoidal curve, which is itself presumed to better approximate underlying biological/ecological reality. The conceptual basis of my analysis is encapsulated in the following figure:



Conceptual Argument

The first portion of the sigmoidal curve is convex. In this zone, log RGR is a decent approximation to underyling growth. However, there is an adjcent, larger zone of approximate linearity, where the linear growth rate describes biological reality more closely. Finally, neither approximation is great in the upper portion of the curve, well into its concave portion, although the linear approximation is uniformly superior throughout the concave zone.

### Mathematical Analysis of Log RGR

## General Taylor Series Approximation

Mathematically, the argument can be boiled down for any generic equation for growth over time: . Using Taylor Series, we can approximate around some value to second order with a generic function as:

As noted above, the canonical log RGR corresponds to exponential dynamics . Use of exponential dynamics to approximate obviously only works well where both the first and second derivative of are positive (i.e. where function is convex). Given that the second derivative of any sigmoidal curve flips from positive to negative, this approximation error grows rapidly outside of a narrow zone.

## Analysis

We take the familiar logistic equation as a reasonable representaton for sigmoidal growth, while noting that many options are available:

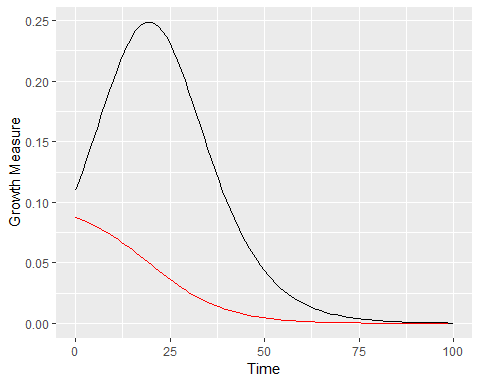
Given this representation of underlying growth, the first question is: what does the canonical log measure correspond to? In other words, we want to ask what happens given exact measurements of and , given that they are sampled from above equation. For We have:

If we have size observations and from two times, and ,the difference between them is:

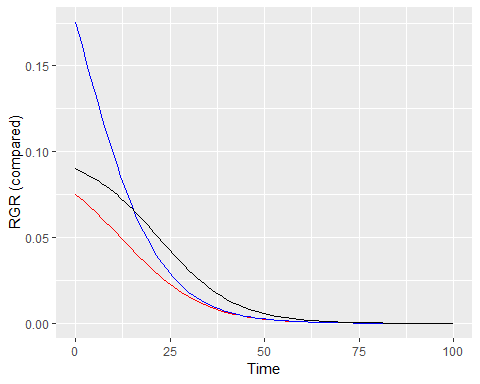
For any given interval :

One flaw of log RGR (as pointed out previously by XXXX) is that RGR is really time-varying, but in effect treated as though time constant (by necessity given the limitation of data). If we want to investigate how this quantity varies with sampling of arbitrary timepoints and along the sigmoidal curve, we re-express and let

### Log RGR versus Linear RGR

Graphically, the comparison with the observed growth increments is: 

Although striking, this comparison is misleading since the quantites differ in dimesion. The log measure is really a **rate** with dimensions of , and I should note that the usual re-expression as is unhelpful at best. In order to make a dimensionally valid comparison, the linear measure must be divided by the initial mass. Here is the result of doing so:



As expected, the log RGR outperforms the linear RGR in matching the real (time varying) RGR early in the convex portion of the curve, then gets outperfomed by the linear RGR thereafter. Both become pretty bad in the concave portion of the curve.

The question is, given a set of data with one or two observation periods, what quantity should be analyzed? The log RGR, the linear growth rate, or perhaps the linear RGR? First, as noted above, these quantities are answers to different questions at some level. The RGR measures differ in dimension from the linear growth rate measure. The answer the question: how much new growth occurs as a function of size? The linear growth rate simply answers: how much new growth has occurred?

**The only reason to prefer the former is the idea that, in the long run, it will better predict growth dynamics.** We have already rejected the idea that, in the data limited setting considered here, we are getting an accurate understanding of time-varying growth. In order to game this out, we need to look at implied dynamics.

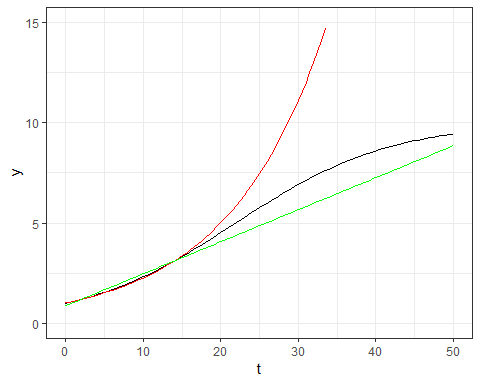
### Comparing log RGR and linear growth in terms of dynamics

Use of the linear growth rate corresponds to assumption of a static linear growth rate dynamic, just as use of corresponds to assuming a static exponential growth rate dynamic. In the latter case, the log RGR has the nice property of representing an ergodic observable (sensu Peters and Gell-man 2016), but is only truly valid assuming exponential growth. While widely (and rightly) dismissed as unrealistic, the linear growth rate may in fact be a generally superior measure for ecological analysis where no time series of size/biomass data is available.

The comparison made here is the goodness of fit implied by replacing the sigmoidal with either a linear approximation or an exponential approximation, given sampling of size from two pairs of time points: 1) from the early (“exponential”) portion of sigmoid curve, and 2) from the middle (“linear”) portion of sigmoidal curve, and 3) from the saturating part of curve.

1. and :

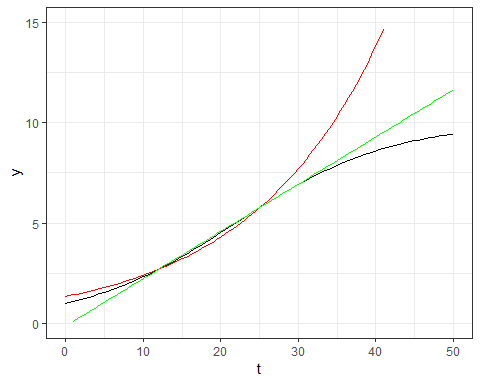
## Warning: Removed 33 rows containing missing values (geom\_path).

 As expected, the exponential approximation works better with data from the convex portion of growth curve. However, the improvement is marginal in absolute value, and quickly diverges outside of the convex portion (in accord with our intuitive model).

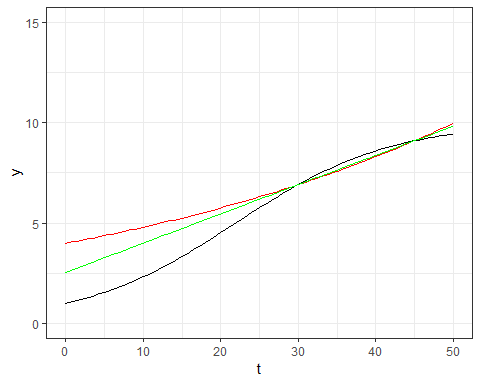
1. and :

## Warning: Removed 18 rows containing missing values (geom\_path).

## Warning: Removed 2 rows containing missing values (geom\_path).

 As can be seen, the linear model is a better approximation where data are taken from within the center part of the growth cycle. Again, the improvement is marginal, but real. Forecast accuracy is much higher, and backcast accuracy marginally worse.

1. and

 As expected, in this scenario, the linear approximation is uniformly better and thus always to be preferred.

At this point, we have seen that the linear growth rate has a large range in which it is superior to the log RGR in terms of implied dynamics (ability to forecast real growth), as well as in terms of approximating the “true” instantaneous RGR.

### Statistical Properties of Log RGR

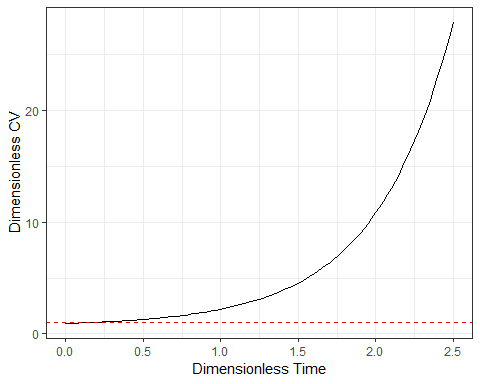
We can derive the sampling distribution of the log RGR based on a Taylor Series’ approximation. Specifically, we consider measurements and with normally distributed error, where the variance scales with the mean (a fairly typical property in biological/ecological data). The distribution of - is then simply the difference of two Normals. Next, we approximate the moments of the distribution of as:

. Using the delta method for variance, we have:

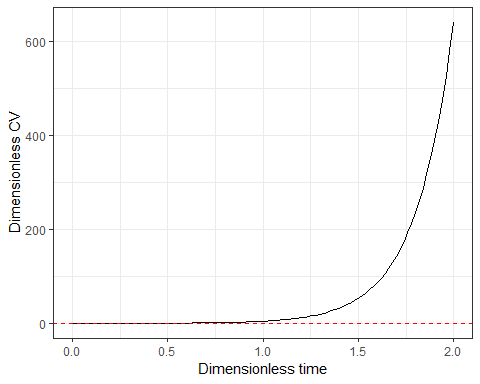
Given a constant coefficient of variation (reflecting variance scaling with mean on the original scale) the sampling distribution of (setting to unit scale), is therefore:

The CV of the log measure is then related to the expectaiton of the Z scores of the new sampling distribution, and is inversely proportional to statistical power:

Thus, where over a unit time increment, the log measure should have greater statistical power, while it loses statistical power as the log measure declines (which as we saw in section XXX above occurs faster than the linear difference of course). This can be reformulated as , revealing a *scale free* property of statistical analysis of the log measure. Specifically, wherever the multiplicative growth increase on a unit time scale , the log measure has worse statistical properties than the linear difference.

For our previous parameter value simulations above, here is the curve of with time , expressed in multiples of : 

As can be seen, in this situation, it is always worse! What happens if we accelerate growth rate considerably (10X)



Expressed on a dimensionless time scale representing multiples of we see that there is a small zone of equivalence. But essentially, it is uniformly less powerful. In the end, it is perhaps not too surprising that it is more difficult to estimate an RGR than a GR. Nevertheless, this should give pause to default use of log RGR.

### Case Study

Maybe not necessary here. Theoretical demonstration feels complete. OTOH, data are what motivate this!

### Conclusions and Recommendations

In summary, the chief virtue of the log RGR measure in the data limited setting is that it is best suited to estimate RGR in approximately exponential phases of growth. In a narrow band of the sigmoid curve, it is arguably superior to working with linear growth rate or RGR.

The much maligned linear growth rate is a superior default on three grounds therefore. First, it represents an intuitive quantity with dimensions of size/length/mass, whereas any RGR has dimensions of ). Second, interpreted in terms of dynamics, it is *a better approximation* than the exponential dynamics implied by the log RGR in many parts of the growth curve. Third, it has superior statistical properties almost everywhere.

The widespread use of the log RGR as an *a priori* preferred default for the data-limited situation should be abandoned. Where only two or three time points are available, fitting a linear growth trend is just as good if not better than estimating an exponential growth rate. Unfortunately, none of the measures discussed here can overcome a critical problem in this setting: we do NOT know what portion of the sigmoidal growth curve we are sampling from, in many cases. In this context, careful thought and attention is needed in making comparisons among possible measures.

The ideal scenario is to collect a proper time series (5-7+) and fit a proper growth model. Where data are at all limiting, I recommend careful incorporation of literature values and other external information as priors in a fully Bayesian analysis in order to regularize inferences.