**TagGrowth: an R package for estimating variation in growth among individuals and over time using tagging data**

Alternative #1: Variation in growth among individuals and over time: a case study and simulation experiment involving tagged Antarctic toothfish

Authors: D’Arcy N. Webber1, James T. Thorson2

1Quantifish

1 Saint Michaels Crescent

Kelburn

Wellington 6012

New Zealand

2Fisheries Resource Assessment and Monitoring Division

Northwest Fisheries Science Center

National Marine Fisheries Service

National Oceanic and Atmospheric Administration

2725 Montlake Blvd. East

Seattle, WA 98112

**Abstract**

Words

**Keywords:** Antarctic toothfish; time-varying growth; random effect; von Bertalanffy growth function

## 1 Introduction

Growth in fishes arises from complex behavioral trade-offs for each individual in a population. Growth (measured as an increase in standard length) arises from the allocation of energy between increased size (length) and increased condition (i.e., weight at length). Individuals can also increase their rate of energy acquisition via increased foraging activity, but increased foraging may also cause changes in the risk of predation. Environmental conditions modify the potential for growth, as well as the trade-offs faced by individual fish. For example, increases in water temperature are associated with increased activity levels but also increased energetic demands for maintaining existing body tissue, such that the net effect of changing temperature on growth may vary among individuals. Similarly, changes in survival rates (via fishery harvest) may favor earlier maturation, thus affecting the relative allocation of energy between growth and reproduction.

Many recent studies of captive or wild populations have demonstrated persistent differences in behavioral or phenotypic traits among individuals (termed differences in “personality”). Persistent differences in activity level or tolerance of predation risk (i.e., a tendancy to forage in high vs. low-quality habitat) will likely lead to persistent differences in growth rates among individuals. For example, persistent differences in growth rate, combined with size-selective harvest, will result in older individuals being composed primarily of slow-growing individuals (termed “Rosa Lee’s Phenomenon”). Failure to account for persistent differences in growth rate can therefore lead to biased estimation of average growth rates in wild populations, and population dynamics models are increasingly developed to account for these effects (Taylor and Methot 2013).

In addition to persistent differences in growth rate among individuals, populations may experience synchronous variation in growth rates in a given year. For example, annual variation in water temperature may drive changes in metabolic costs as well as food availability in upwelling systems such as the California Current (Black et al. 2010), and this may cause all individuals to have elevated or depressed growth rates in a given year. Recent meta-analytic work has supported the hypothesis that time-variation in average growth rates is the rule rather than the exception for marine populations (Thorson and Minte-Vera In press). Similarly, many studies have illustrated persistent spatial differences in growth rates (Gertseva et al. 2010), e.g., where populations often have a smaller maximum size in warmer waters.

We therefore hypothesize that most marine populations will have multiple sources of variability in individual growth rates, including:

1. Persistent variation among individuals (i.e., some individuals grow faster or to a larger size than others)
2. Spatial variation (i.e., some areas have different size at age on average than others)
3. Temporal variation (i.e., all individuals having elevated or depressed growth rates in a given year)
4. Residual variation (i.e., individuals will have spurts and drops in growth rates).

Mixed-effects models are generally used to decompose variance in a data set into multiple sources of variability (Thorson and Minto In press). Mixed-effects models accomplish this by estimating true size at age as a latent variable, which is integrated across during estimation of growth parameters. Hierarchical models are increasingly advocated as a strategy to partition variability into multiple sources, and hence to interpret which source of variance is worth further study (Larsen et al. 2001, Gelman 2005).

Growth rates in natural populations are generally estimated using capture-mark-recapture (CMR) models. In CMR studies for marine fishes, thousands of individuals are typically captured and marked with a persistent and unique tag, and then measured and released. If a tagged individual is recaptured, it is often killed and its age at recapture identified via analysis of hard parts. Given the known time between capture and recapture, its size and age at both times can be calculated. This study design allows a model for growth increments to be fitted to size and age data for two times for each individual in a data set.

Many studies have examined the impact of persistent or residual variation on estimation of growth patterns using CMR data for wild marine populations (Sainsbury 1980, Francis 1988). However, few studies have explicitly modeled multiple sources of variability in growth rates. One exception is Shelton et al. (2013), which modeled persistent, temporal, and residual variation in growth rates in the analysis of data from a growth experiment for steelhead trout. In this study, we additionally incorporate sex-specific differences and spatial variation in growth rates, and implement the model using maximum marginal likelihood estimation. The code is distributed as a publicly available package in the R statistical environment, and is illustrated using CMR data for Antarctic toothfish (*Dissostichus mawsoni*). Using these data, we show that …. [1-2 sentence summary of most important result for toothfish].

**2 Methods**

We start with the specialized von Bertalanffy growth function:

(1)

where *dL/dt* is change in length *L* as a function of time *t*, *a* scales with energy acquisition, and *k* represents metabolic upkeep costs. Variation among individuals can be estimated by treating each individual's demographic parameters as a random effect that arises from a population-level distribution (Thorson and Minto In press). Here, we treat metabolic upkeep costs *ki* as varying among individuals, due to differences in ambient temperature, intrinsic variation in activity levels, and other factors:

where , *μk*  is the median *ki* among individuals in the population, and *σk* is the standard deviation of individual upkeep cost deviations .

Individuals that are more highly active may obtain more food (increased *a*) and simultaneously have greater upkeep costs (increased *k*). Following Shelton et al. (2013), we include this association via the following equation:

(3)

where *Ψ* determines whether upkeep costs (*ki*) and energy acquisition rate (*ai*) are highly correlated (*Ψ=1*) or independent among individuals (*Ψ=0*), and *γ* controls the average energy acquisition rate among individuals.

We also assume that environmental conditions experienced by each individual will vary over time, such that each individual will have variation over time in their upkeep costs (*ki*), in turn affecting their access to food (*ai*). In particular, we assume that upkeep costs for a given individual *i* for a short time interval *t* can be approximated by a constant value (*ki,t*), which will vary around the average value for this individual (*ki*). In the proceeding, we use a time interval of one week (7 days), while confirming that results are similar for other small time intervals. Integration then yields:

(4)

where

where *Δt* is the number of time-periods (weeks) elapsed between length intervals and *nΔ* is the number of time-periods per year (52.15), such that *k* is the Brody growth coefficient (with units years-1). We assume that all age-0 individuals have a length *L0,s* that differs between males *(s=0*) and females *(s=1*), and this age-at-birth for each sex is estimated. Eq. 4 can then be used to calculate the predicted length at of first capture *L(t1)* and upon recapture *L(t1+t2*), where *t2* is the time between first tagging and recapture, and *t1+t2* is the age at recapture as determined from conventional ageing methods.

observed length for each individual upon first tagging *Lobs(t1)* and upon recapture *Lobs(t1+t2*) is assumed to follow a normal distribution with a fixed coefficient of variation (representing observation error caused by at-sea length measurements) given the estimated length at that age *L(t)*

(5)

, and

## 2.1 Estimation

We estimate the value of fixed effects in this model (, , , ,) using maximum marginal likelihood while marginalizing across the probability of all random variables representing variation among individuals and over time (Thorson and Minto In press). Preliminary exploration using data for the case study (explained below) suggested that *Ψ* was not estimable, so we proceed by fixing *Ψ=0*. Parameter estimation is conducted using Template Model Builder (TMB; Kristensen (2014);available from: <https://github.com/kaskr/adcomp>) software called from the R statistical environment (R Core Development Team 2013) and all code necessary for replicating the case study and simulation experiment are publicly available (<https://github.com/quantifish/TagGrowth>, Appendix A). Estimation was conducted. The marginal likelihood is then maximized using conventional nonlinear minimization tools in R, while using the gradient of the marginal likelihood with respect to fixed effects as calculated by Template Model Builder. Standard errors are then estimated via the information matrix and delta-method.

After fixed effects have been estimated, the median asymptotic length for individuals in the population can be calculated as:

(6)

which is obtained by setting Eqn. 1 equal to zero, substituting in Eqn. 3, and rearranging. Readers are referred to Shelton et al. (2013) for an expanded model that also incorporates variability in average energy acquisition rates for all individuals in a given time period.

**2.2 Case study application**

Antarctic toothfish are large Nototheniids native to the Southern Ocean. They can grow to be more than 2 m in length, weighing over 100 kg, and can live for up to 50 years of age. The exploratory toothfish fishery in the Ross Sea region began in 1997 and is managed by The Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR). Since then the fishery has increased to about 3000 tonnes per year. Fishing is restricted to the summer months (months), once the ice shelf recedes allowing vessels access to the region. The Antarctic toothfish tagging programme was initiated in the in the 2001 fishing season by New Zealand vessels involved in the fishery. In 2004, toothfish tagging was made compulsory for all vessels participating in the fishery. The tagging programme records information on the date, depth, location, sex, and size of each tagged/recaptured fish. A small subset of the recaptured fish are aged by reading their otolith.

We identified those fish that had been tagged, recaptured and aged upon recapture. This yielded 315 individuals of which 166 were female and 149 male. We fit several different versions of the model to these data, these included:

1. No random-effects
2. Random-effects for upkeep costs only (log(*ki*))
3. Time varying individual random-effects only (*zi,t*)
4. Time varying individual random-effects and random-effects for upkeep costs (log(*ki*) and *zi,t*)
5. Same as 4 above but include fishing year effect

The simulation study illustrated that the model had little ability to identify the value of *Ψ*. We therefore fixed the value of *Ψ*=0. We also fit a model where fishing year *y* (a fishing year was defined as 1 Nov. to 31 Oct., *t* ⊆ *y*) is incorporated as a random-effect applied to the average energy acquisition rate parameter *γ*, thus we have:

where (6)

This requires the estimation of the median energy acquisition rate parameter *μγ* and a deviation parameter from the fishing year that the first fish was born (1972) until the most recent fishing year that a fish was recaptured (2011). Preliminary exploration indicated that seasonal effects could not be explored because all recaptures are observed during a few months (X-Y) and hence seasonal and annual effects are confounded.

## 2.1 Simulation

A simulation study was done to ensure that the model could generate estimates of growth rates that are unbiased and reasonably precise. Four different scenarios were simulated including: no random-effects, random-effects for the upkeep costs *ki* only, time varying random-effects for *z* only, and random-effects for *ki* and *z*. A power analysis was done for each of the four scenarios where data was available for 50, 100, 250 and 500 simulated individuals. This yielded 16 different simulation experiments. We did 200 replicates for each of the 16 combination of scenario and sample size.

The model parameters *t0*, *k*, *L∞* and the c.v. of growth were taken from Dunn et al. (2006) and used to derive the values of several parameters in the simulation study. Where required parameters were converted to have weekly units (i.e. *k*, *t0*, age). The parameters *L0* and *γ* can be calculated as:

(7)

and

(8)

The parameters *L0*, *k* and *γ* were all set to be sex-specific (Table 2a). Initial exploration illustrated that the model had little ability to identify the value of *Ψ*. We therefore fixed the value of *Ψ*=0. The c.v. of observed growth *cobs* was set at the value in Dunn et al. (2006) in the scenario with no random-effects. In the remaining scenarios, the c.v. and the standard deviation parameters, *σk* and *σz*, were set at values that resulted in reasonable variation in individual length trajectories (Table 2b). The parameter *σk* was set to be sex-specific, *c*obs and *σz* were not.

For each individual in each of the scenario/power simulations, sex was sampled with replacement from the observed sex structure in the toothfish data set. The age at release, age at recapture and time at liberty were sampled independently from the observed toothfish data set also, then one of the three variables was chosen at random and calculated using the remaining two variables (e.g., if age at recapture was chosen at random it was calculated as the age at release plus the time at liberty) and rounded to the nearest integer.

The model was fit to the data for each of the 200 replicates in each of the 16 simulation experiments.

## 3 Results

## 1

## 3.2 Simulation

* Do table showing number of pdh fits for each of the 4 scenarios. Also show true par values, and mean/median of estimates. Discuss the numbers of pdh fits for each simulation experiment.
* Parameters *μk*, gamma and *L0* well estimated in all cases, distribution of estimates gets tighter as the power is increased from 50 to 500. Figure X.

## 4 Discussion

In this study, we have used a case study application for Antarctic toothfish and a simulation experiment to demonstrate that it is feasible to simultaneously estimate variation in growth among individuals and years, between sexes, and also including residual variation in growth rates for each individual over time, in addition to measurement errors. Early studies incorporated variability in growth via individual variation in growth parameters (Sainsbury 1980) or both time-dependent and time-independent variation in growth increments (Francis 1988). Recent studies have developed mixed-effect models for estimating variation in growth rates over time and/or among individuals in wild or experimental settings involving fish (Stawitz et al. In review, Thorson and Minte-Vera In press, Shelton et al. 2013, Vincenzi et al. 2014). Variation in growth rates is also an active area of research and method development in other ecological taxa (Brooks et al. 2013). However, our study is the first to our knowledge to apply these methods to a wild marine population. Our case study application suggests that … [something about relative size of estimated SDs]

By estimating variation among individuals, over time (both shared and differing among individuals), and between sexes, we have incorporated most types of growth variation that would be expected in a wild population. However, we have neglected one obvious type of variation in growth rates – variation over space. Previous research has demonstrated the prevalence of individual variation in growth (Gertseva et al. 2010) and condition (Thorson In press) over space. The modelling framework we present here could easily be modified to include spatial variation in growth rates, either among spatial strata or using geostatistical techniques (Thorson et al. In press, Shelton et al. 2014), but in this study we have chosen instead to focus on testing the individual, temporal, and sex-specific variation that we presented here. We hypothesize that including spatial variation would decrease the magnitude of among-individual variation in growth rates, because spatial variation would explain some portion of individual-level variation.

We have also distributed our new software tool as an R package, and hope that easy access to computational tools will allow comparison of results among species, regions, and taxa. There is little existing research regarding the relative magnitude of individual, temporal, and spatial forms of variation in growth. This information is important for at least two reasons. First, stock assessment models and fisheries management strategies are routinely evaluated using simulation experiments (termed “management strategy evaluation”; Sainsbury et al. (2000)). However, simulation experiments depend upon having information regarding plausible states-of-nature, which are used to simulate available data (Punt 2008). Models estimating time-variation in growth can be used to simulate more plausible data for use in testing assessment methods. Second, previous research has indicated that growth estimates are biased when variation when estimated via models that do not account for variation in growth (Sainsbury 1980), and that these biases can in turn result in biased estimates of stock status and productivity (Punt et al. In review).

* Mention how when the model was fit to annual rather than daily growth increments that the estimated parameters resulted in biased high growth schedules.
* Growth shock or growth retardation due to tagging. Tag shock has been estimated before as about 0.7 year lost in growth. Confounded by live/dead.
* Snapper may also have some data like this.
* Seasonality

**Acknowledgements**

We thank in particular O. Shelton for developing and discussing the model for variation in growth that was used here. We also thank K. Kristensen and H. Skaug for developing and maintaining the Template Model Builder software, which facilitates this study and many others.

**Bibliography**

Black, B.A., Schroeder, I.D., Sydeman, W.J., Bograd, S.J., and Lawson, P.W. 2010. Wintertime ocean conditions synchronize rockfish growth and seabird reproduction in the central California Current ecosystem. Can. J. Fish. Aquat. Sci. **67**(7): 1149–1158.

Brooks, M.E., McCoy, M.W., and Bolker, B.M. 2013. A Method for Detecting Positive Growth Autocorrelation without Marking Individuals. PLoS ONE **8**(10): e76389. doi: 10.1371/journal.pone.0076389.

Francis, R. 1988. Maximum likelihood estimation of growth and growth variability from tagging data. N. Z. J. Mar. Freshw. Res. **22**(1): 43–51. [accessed 19 December 2014].

Gelman, A. 2005. Analysis of variance—why it is more important than ever. Ann. Stat. **33**(1): 1–53.

Gertseva, V.V., Cope, J.M., and Matson, S.E. 2010. Growth variability in the splitnose rockfish Sebastes diploproa of the northeast Pacific Ocean: pattern revisited. Mar. Ecol. Prog. Ser. **413**: 125–136. [accessed 3 May 2013].

Kristensen, K. 2014. TMB: General random effect model builder tool inspired by ADMB. Available from https://github.com/kaskr/adcomp.

Larsen, D.P., Kincaid, T.M., Jacobs, S.E., and Urquhart, N.S. 2001. Designs for Evaluating Local and Regional Scale Trends. BioScience **51**(12): 1069–1078. doi: 10.1641/0006-3568(2001)051[1069:DFELAR]2.0.CO;2.

Punt, A.E. 2008. Refocusing stock assessment in support of policy evaluation. *In* Fisheries for Global Welfare and Environment. *Edited by* K. Tsukamoto, T. Kawamura, T. Takeuchi, T.D. Beard, and M.J. Kaiser. TerraPub, Tokyo. pp. 139–152.

Punt, A.E., Haddon, M., and McGarvey, R. In review. Estimating Growth within size-structured fishery stock assessments: What is the state of the art and what does the future look like?

R Core Development Team. 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available from http://www.R-project.org/.

Sainsbury, K.J. 1980. Effect of individual variability on the von Bertalanffy growth equation. Can. J. Fish. Aquat. Sci. **37**(2): 241–247. [accessed 19 December 2014].

Sainsbury, K.J., Punt, A.E., and Smith, A.D.. 2000. Design of operational management strategies for achieving fishery ecosystem objectives. ICES J. Mar. Sci. J. Cons. **57**(3): 731–741.

Shelton, A.O., Satterthwaite, W.H., Beakes, M.P., Munch, S.B., Sogard, S.M., and Mangel, M. 2013. Separating intrinsic and environmental contributions to growth and their population consequences. Am. Nat. **181**(6): 799–814. [accessed 8 October 2013].

Shelton, A.O., Thorson, J.T., Ward, E.J., and Feist, B.E. 2014. Spatial semiparametric models improve estimates of species abundance and distribution. Can. J. Fish. Aquat. Sci. **71**(11): 1655–1666. doi: 10.1139/cjfas-2013-0508.

Stawitz, C.C., Essington, T.E., Branch, T.A., Haltuch, M.A., Hollowed, A.B., and Spencer, P.D. In review. A state-space approach for measuring growth variation and application to North Pacific groundfish.

Taylor, I.G., and Methot, R.D. 2013. Hiding or dead? A computationally efficient model of selective fisheries mortality. Fish. Res. **142**: 75–85. [accessed 20 April 2013].

Thorson, J.T. In press. Spatio-temporal variation in fish condition is not consistently explained by density, temperature, or season for Northeast Pacific groundfishes.

Thorson, J.T., and Minte-Vera, C. In press. Relative magnitude of cohort, age, and year effects on size at age of exploited marine fishes. Fish. Res.

Thorson, J.T., and Minto, C. In press. Mixed effects: a unifying framework for modelling in aquatic ecology. ICES J. Mar. Sci.

Thorson, J.T., Skaug, H., Kristensen, K., Shelton, A.O., Ward, E.J., Harms, J., and Benante, J. In press. The importance of spatial models for estimating the strength of density dependence. Ecology. doi: http://dx.doi.org/10.1890/14-0739.1.

Vincenzi, S., Mangel, M., Crivelli, A.J., Munch, S., and Skaug, H.J. 2014. Determining Individual Variation in Growth and Its Implication for Life-History and Population Processes Using the Empirical Bayes Method. PLoS Comput Biol **10**(9): e1003828. doi: 10.1371/journal.pcbi.1003828.

Table 2a – Parameter values used for all scenarios in the simulation experiment.

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | **Units** | **Female** | **Male** |
| *L0* | cm | -0.3409 | 3.9777 |
| *k* | weeks-1 | 0.00173 | 0.00179 |
| *γ* |  | 0.3119 | 0.3024 |
| *Ψ* |  | 0.0 | 0.0 |

Table 2b – Parameter values that vary among scenarios in the simulation experiment.

|  |  |  |  |
| --- | --- | --- | --- |
| **Scenario** | **Parameter** | **Female** | **Male** |
| No variation | *cobs* | 0.102 | 0.102 |
|  | *σk* | - | - |
|  | *σz* | - | - |
| Variation in *k* | *cobs* | 0.05 | 0.05 |
| *σk* | 0.1 | 0.2 |
|  | *σz* | - | - |
| Variation in *z* | *cobs* | 0.05 | 0.05 |
| *σk* | - | - |
|  | *σz* | 0.3 | 0.3 |
| Variation in *k* and *z* | *cobs* | 0.05 | 0.05 |
| *σk* | 0.1 | 0.2 |
| *σz* | 0.3 | 0.3 |

Table 1 – List and definition of symbols used in the text and equations.

|  |  |  |
| --- | --- | --- |
| **Parameter name** | **Symbol** | **Units** |
| The number of time periods per year |  | - |
| The hypothetical length at which time | *L0* | cm |
| Anabolic rate | *a* |  |
| Energy expenditure rate (“Brody growth coefficient”) | *ki*, *μk* | years-1 |
| Anabolism coefficient | *γ, γy*, *μγ* |  |
| Allometric scaling of anabolism and energy expenditure | *Ψ* |  |
| Time variation | *z* | cm |
| Asymptotic maximum length |  | cm |
| The hypothetical time at which | *t0* | years |
| Observation coefficient of variation (c.v.) | *cobs* | - |
| Catabolic rate standard deviation | *σk* | cm |
| Annual deviation standard deviation | *σy* | cm |
| Time variation standard deviation | *σz* | cm |
| **Data** | **Symbol** | **Units** |
| Time interval |  | weeks |
| Length at time *t* | *L(t)* | cm |
| Year at time *t* | *Y(t)* | years |
| **Indices** | **Symbol** | **Units** |
| Individual | *i* | - |
| Time | *t* | weeks |
| Sex | *s* | - |
| Year | *y* | years |

Table 2 – List of simulation results.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Random effects | pdH | L0 | k | gamma | Sd obs |  |  |  |
|  |  |  |  |  |  |  |  |  |
| None |  |  |  |  |  |  |  |  |
| k |  |  |  |  |  |  |  |  |
| z |  |  |  |  |  |  |  |  |
| k and z |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

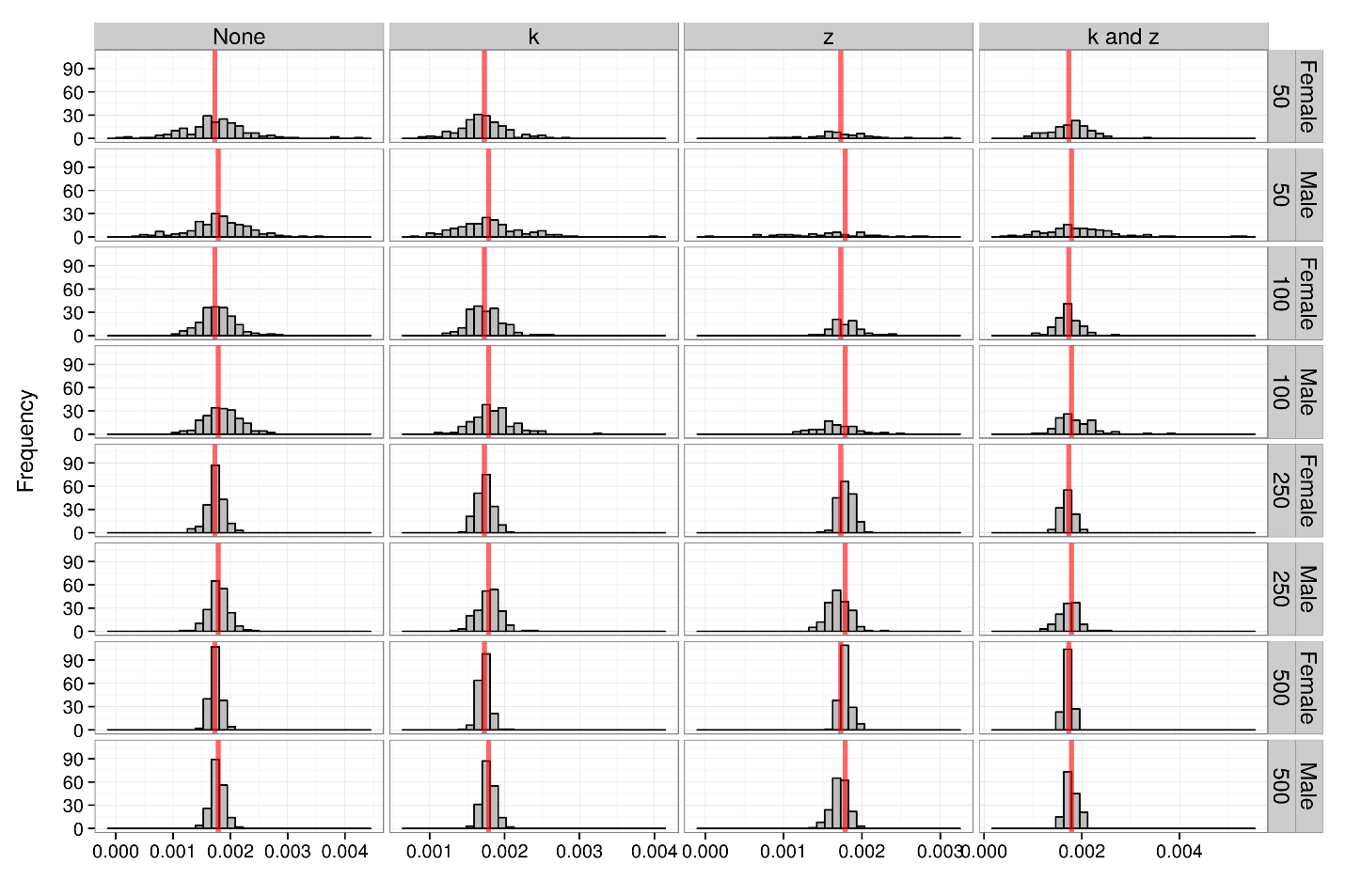


Figure X – Estimated value of muk in each of the simulation study experiments.

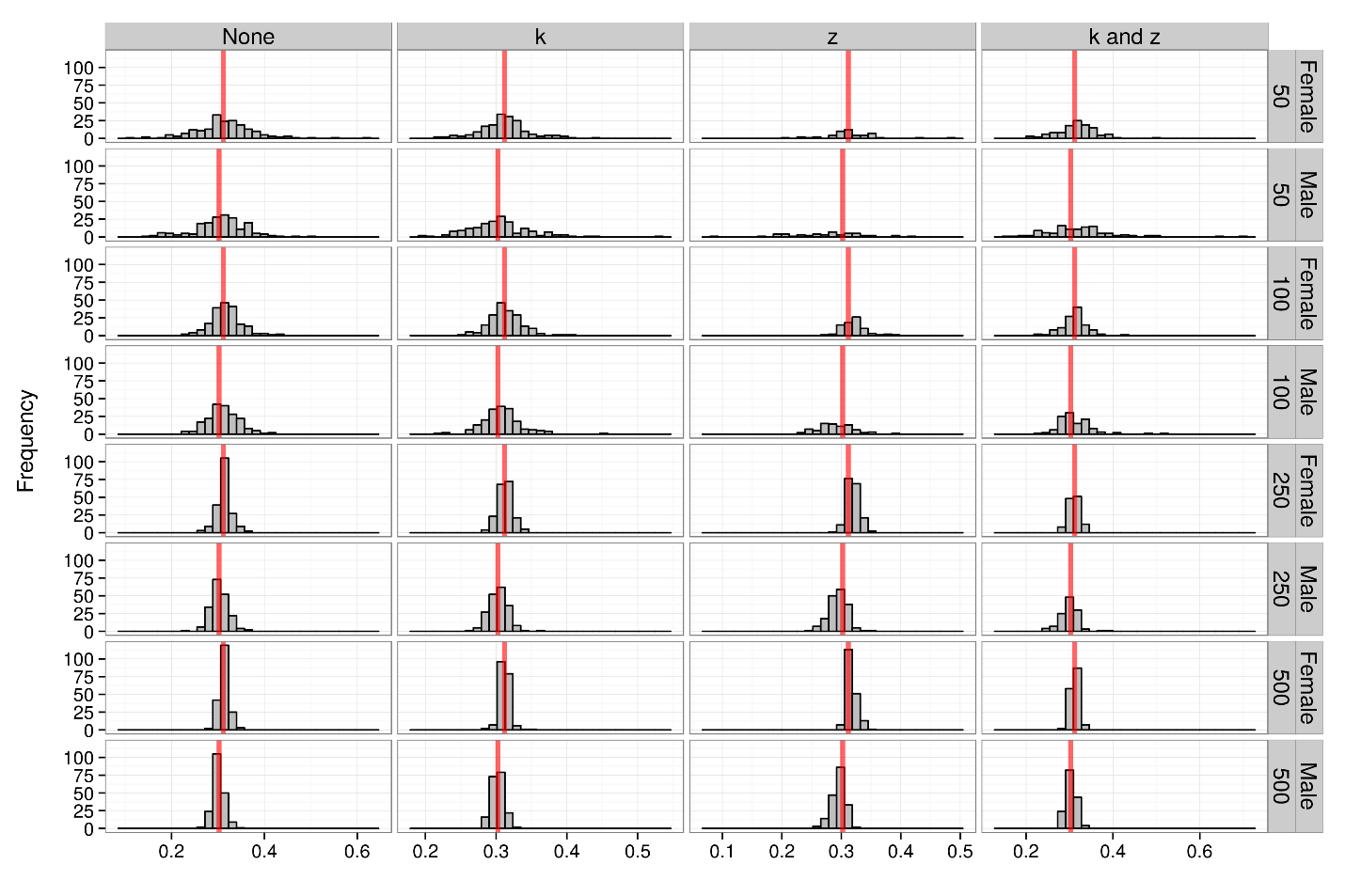


Figure X – gamma sims

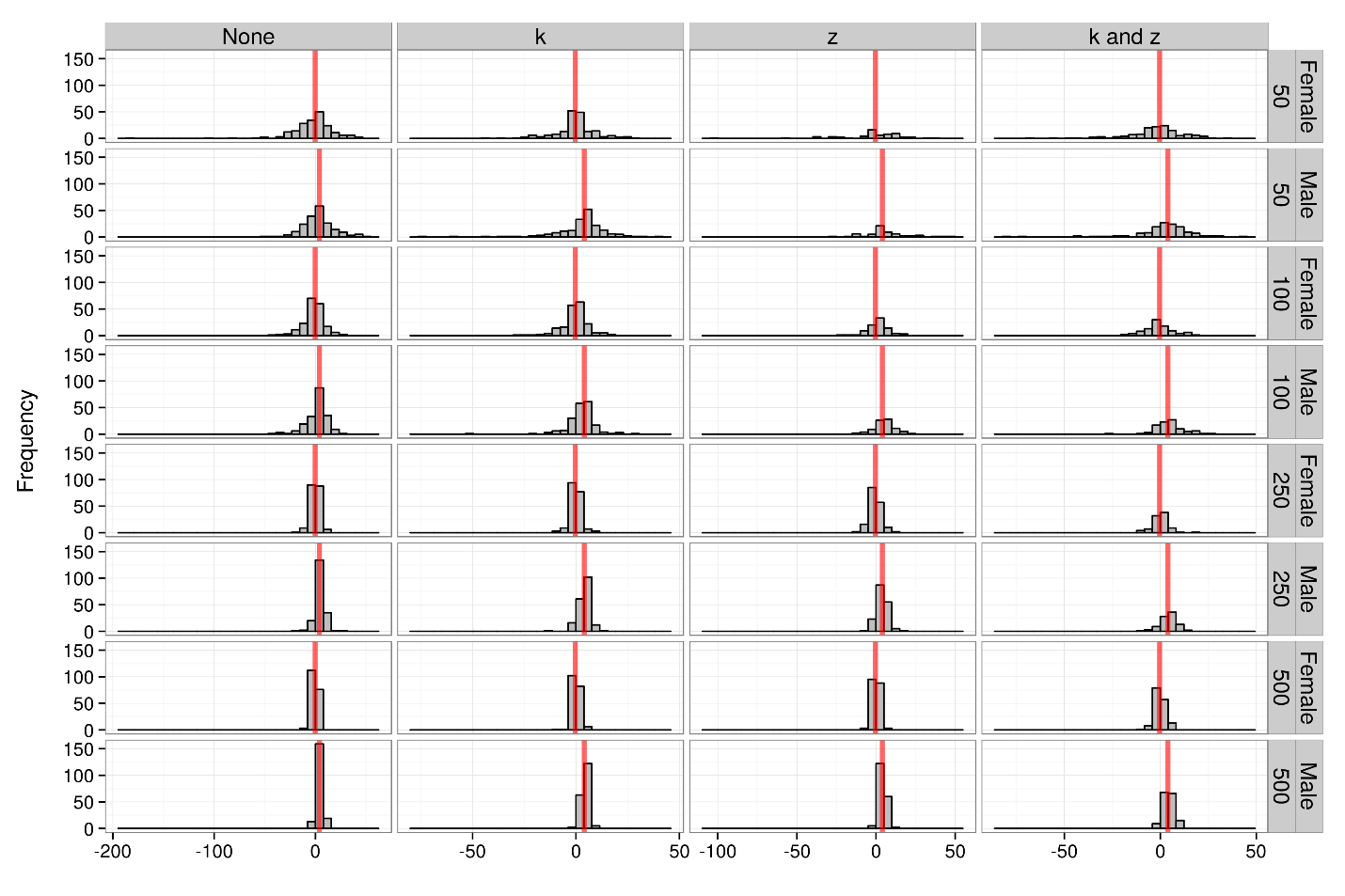


Figure X – L0 sims

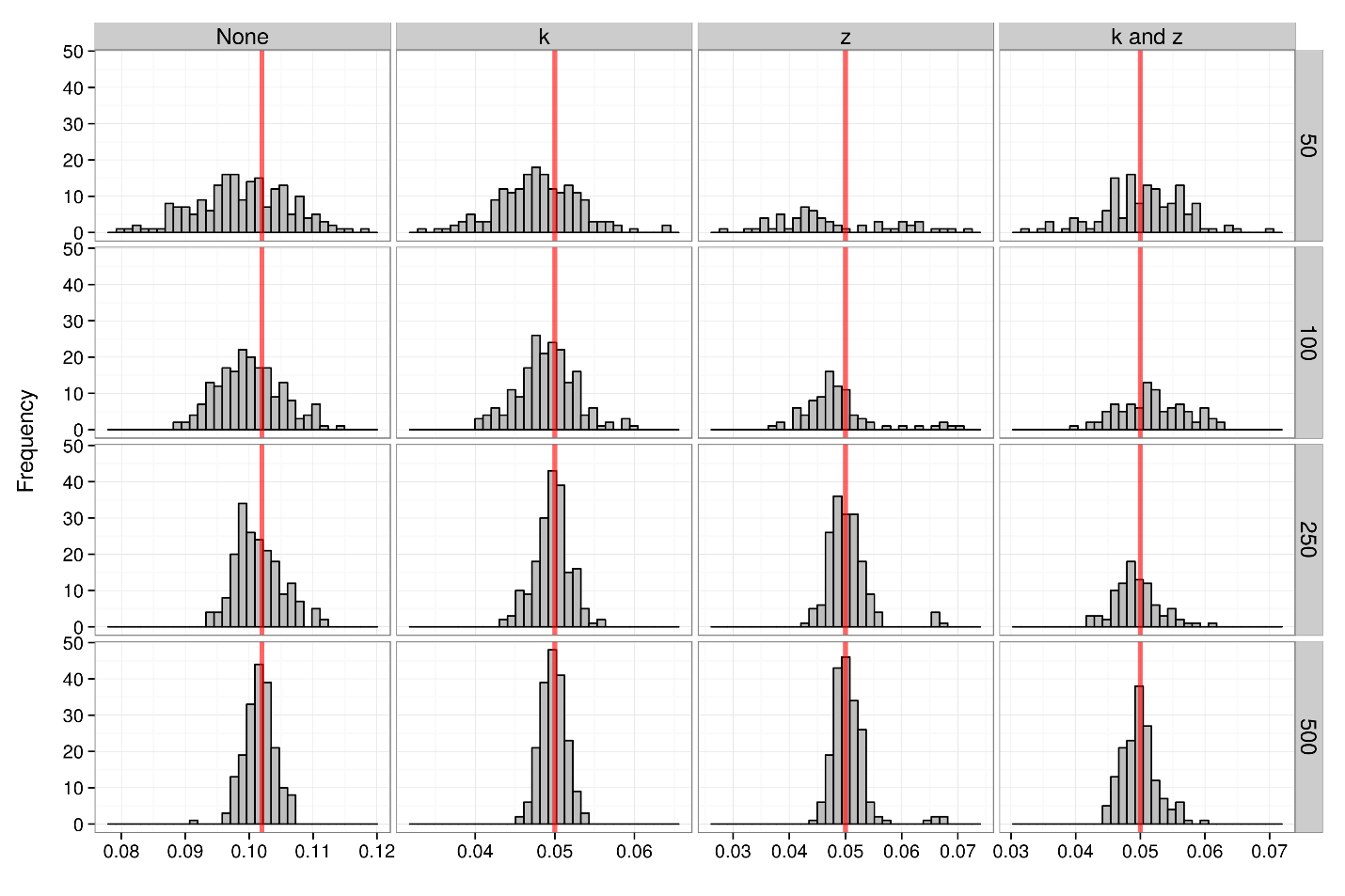


Figure X – cv obs sims

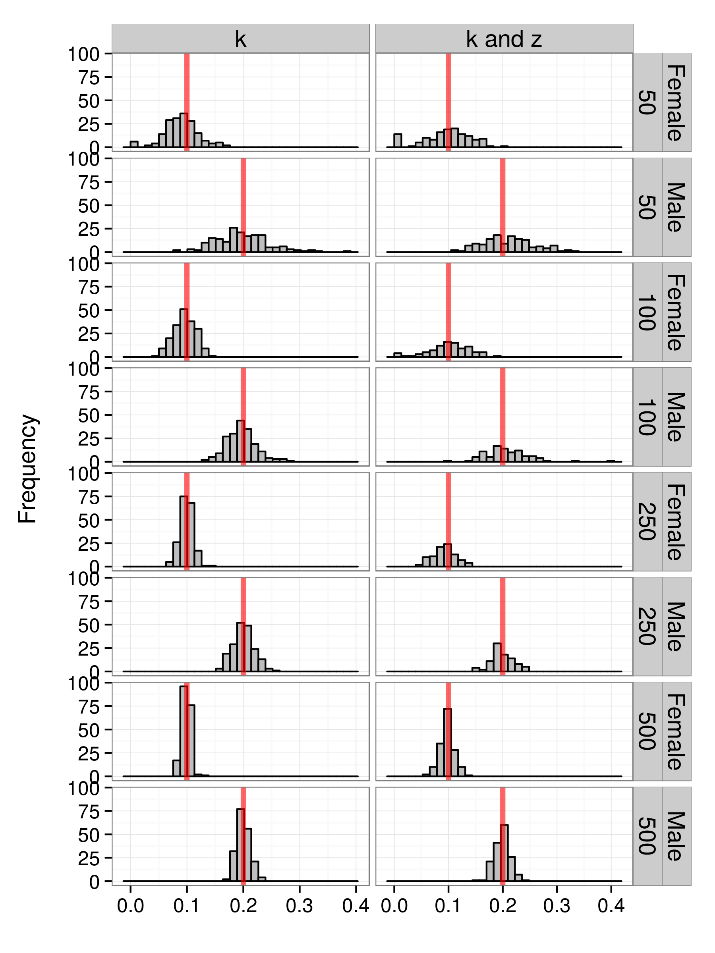


Figure X – kdev sims

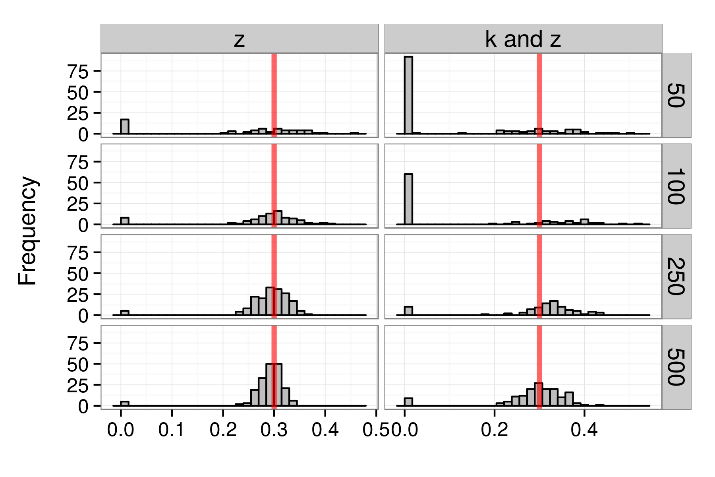


Figure X – sdz sims

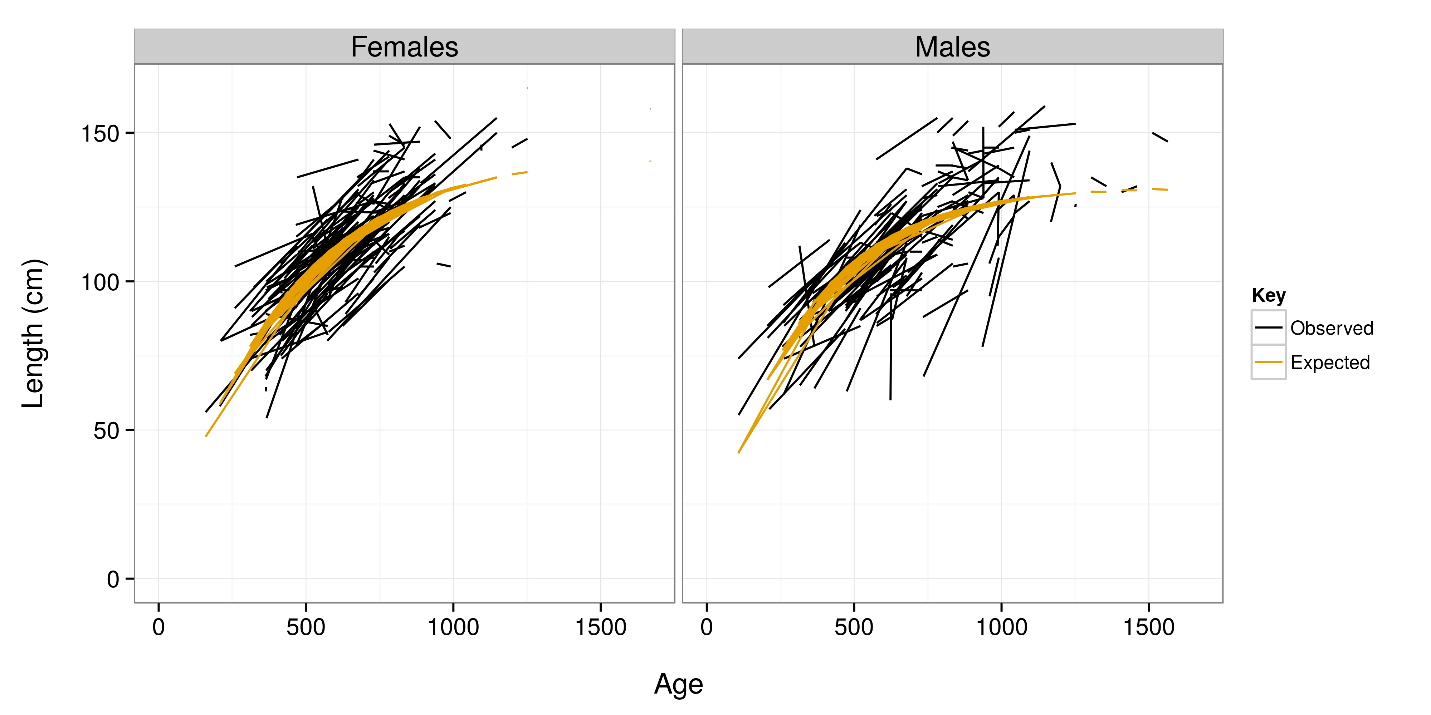


Figure X – no random effects

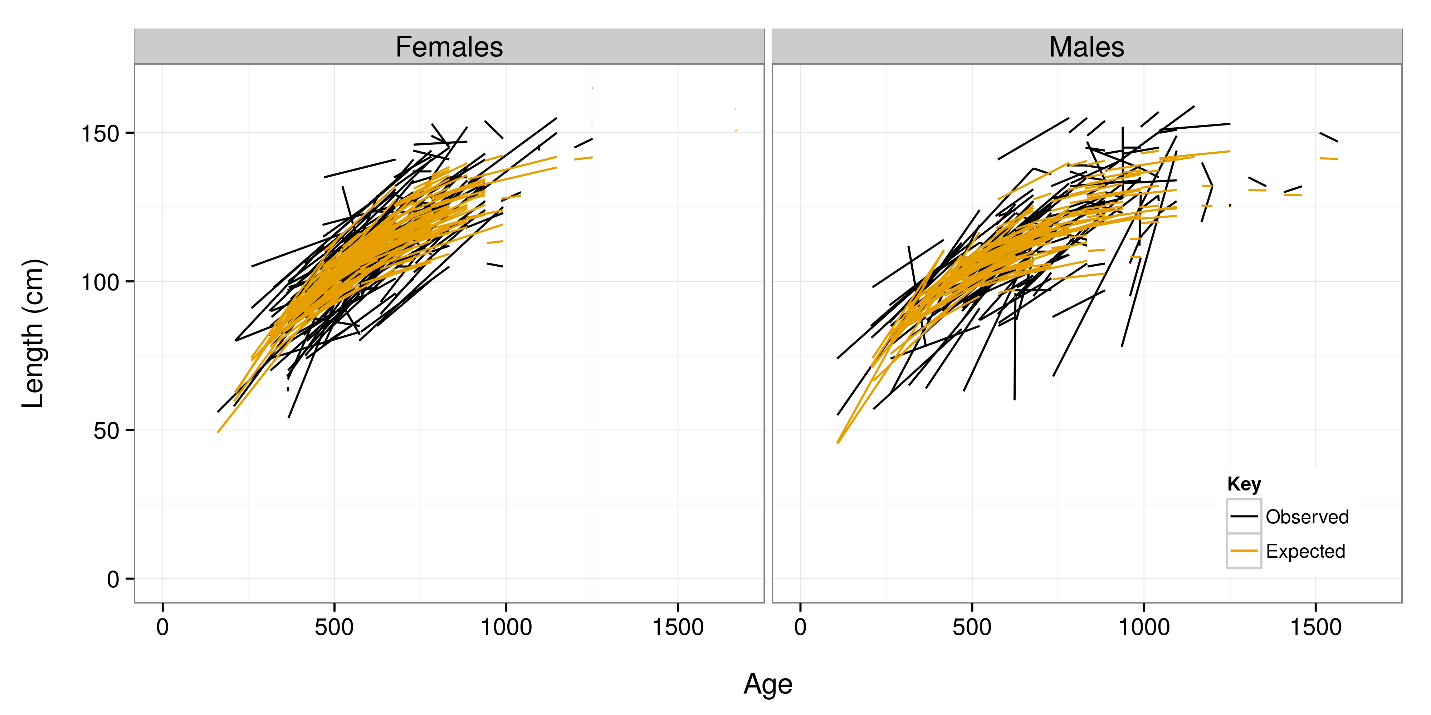


Figure X – random effects of k

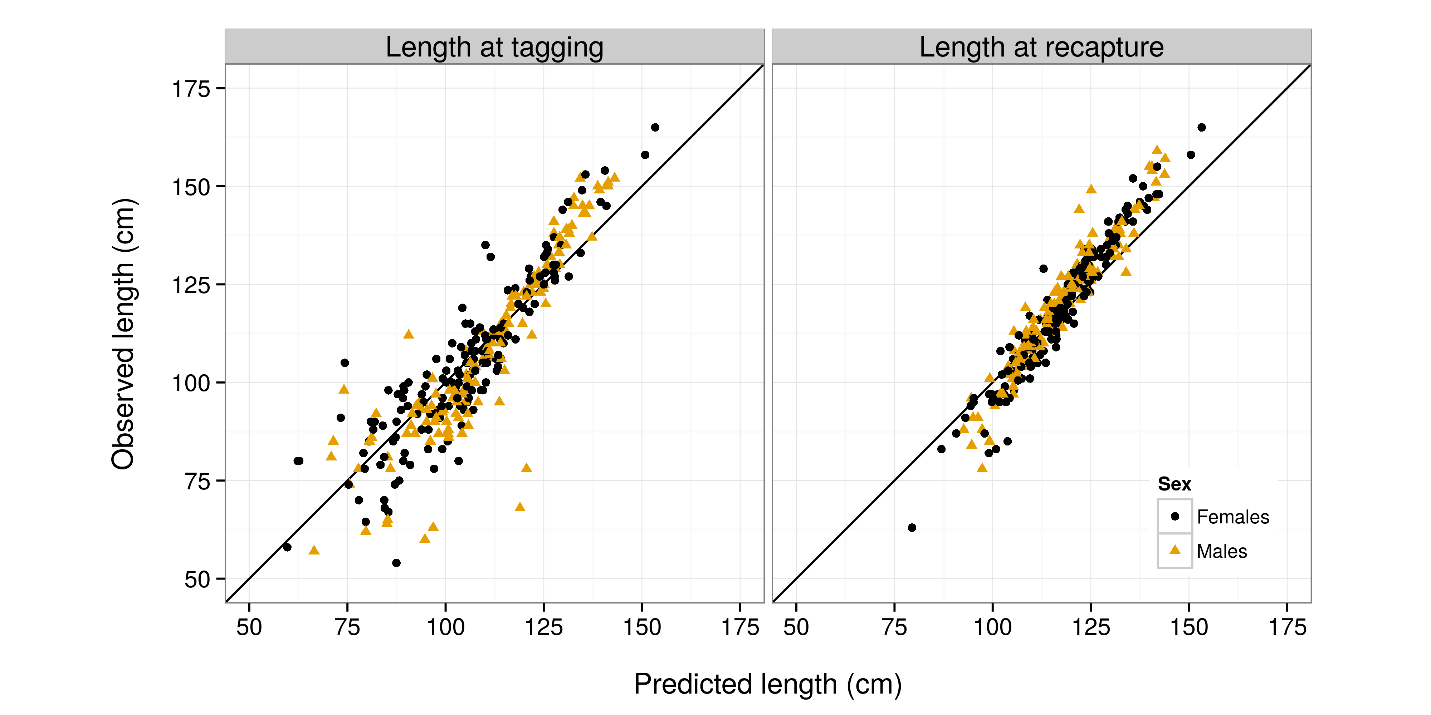


Figure X

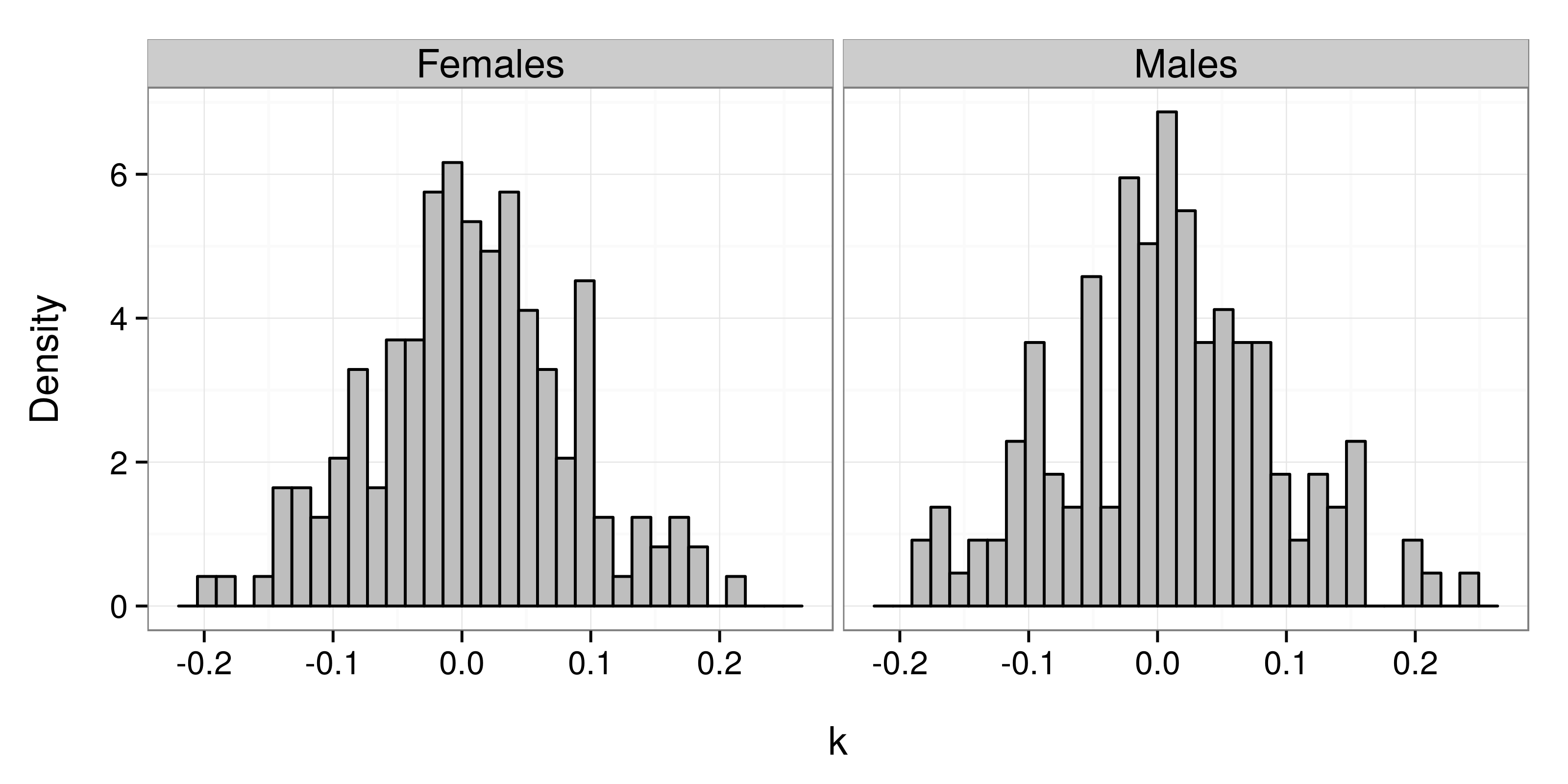


Figure X