Supporting information for: *Test for informative cluster size with right censored survival data*

Alessandra Meddis, Aurélien Latouche

1 Web Appendix A

Simulation study: impact of γ on ICS

We conduct the simulation study to assess the power of the proposed test in various scenarios. We consider clustered survival data characterised by informative cluster size, namely the outcome is depending on the size of the cluster. We generate the data from a frailty model with the frailty term $U_k \sim Gamma(\theta)$ and the cumulative baseline hazard function $A_0(t) = st^{\omega}(s = 6.31e^{-6}, \omega = 4.6)$. To obtain informative cluster size, we consider K clusters with sample size $N_k \sim Pois(\lambda \exp(V_k))$ where λ, common between clusters, represents the expected number of observations in each cluster and V_k defines the cluster-specific sample size. To create the dependence between the sample size N_k and the failure times T_{ik}, we generate (U_k, V_k) from a multivariate Gamma with unit mean and covariance matrix Σ. The variance $\sigma_U^2 = 1/\theta$ defines the variability of failure times among clusters. The variance $\sigma_V^2 = 1/\gamma$ represents the variability between clusters sample sizes. The parameter ρ is the correlation between the two random effects. The strength of ICS depends on θ, ρ, γ.

In this Section we explore how ICS changes with γ. We generate (U_k, V_k) for 100 clusters with $\gamma \in \{3, 10, 40\}$. Figure 1 shows that U_k increases faster for higher γ (less variability), because the range of V_k (sample size) becomes narrower but the range of U_k does not change (θ is fixed). In Figure 2 we provide the mean failure times $\overline{T}_k$ for each cluster sample sizes: for small values of V_k (sample sizes) the U_k will be lower for increasing γ and so failure times will be larger when $\gamma = 40$; bigger values of V_k are associated to bigger U_k and thus to shorter failure times. Therefore, for two fixed sample sizes, the difference of the associated failure times will be larger with an

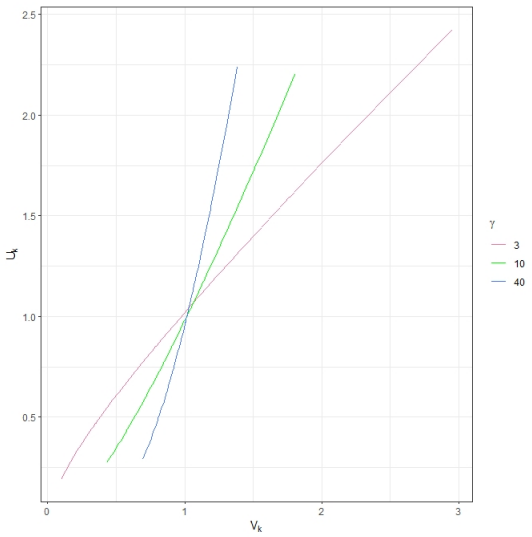

Figure 1: Representation of the two random effects U_k and V_k generated for 100 clusters with different values of γ.

increasing value of γ and informative cluster size is stronger. However, this difference is not visible anymore when the mean clusters sample size decreases ($\lambda = 5$) because the cluster sample sizes are similar ($\gamma = 40$).

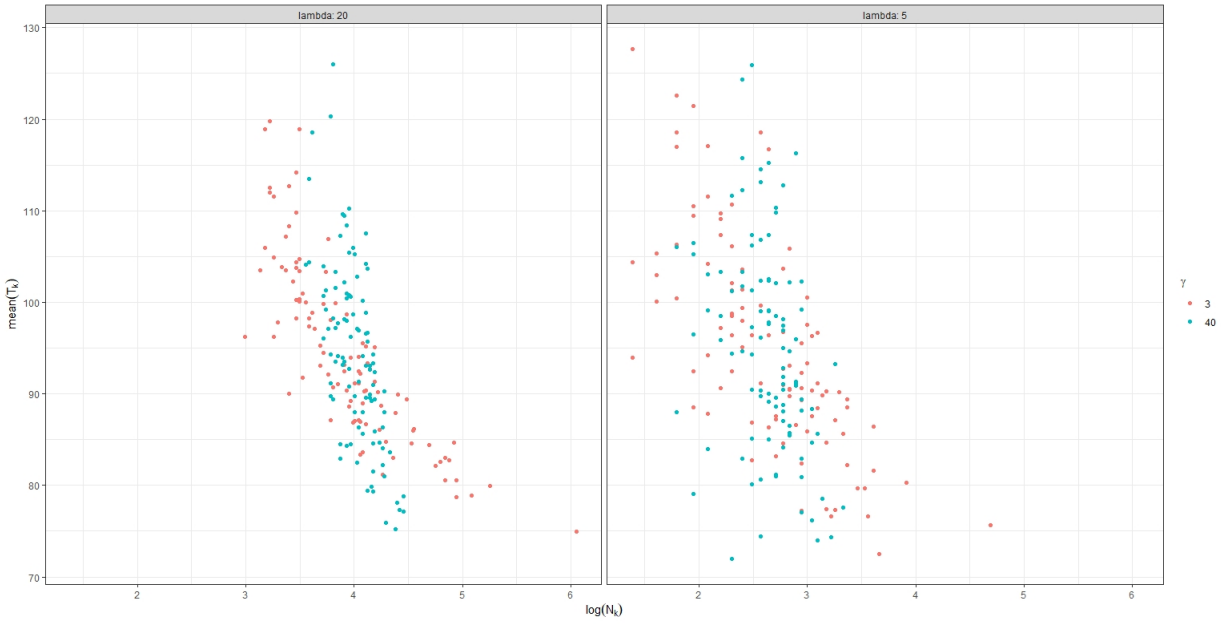

Figure 2: Plot of median failure times T_k and the cluster sample sizes N_k (logarithm scale) associated to the random effects (U_k, V_k). Data for 100 clustered are generated by a shared frailty model and a Poisson distribution as described in the simulation section. The parameter λ of the Poisson distribution represents the mean sample size of clusters if no variability is present in the sample sizes distribution ($\gamma = \infty$).