The Role of Prior Information in Inference on the Annualized Rates of Mass Shootings in the United States
SUPPLEMENTARY MATERIAL

Nathan E. Sanders
Applied Analytics, Legendary Entertainment
and
Victor Lei
Applied Analytics, Legendary Entertainment

March 8, 2018

S1 MCMC simulation

We perform MCMC simulation for the models described in § 3 using the probabilistic programming language *Stan* (Carpenter et al., 2017). We expect the mean function parameters μ_0 and μ_b, the Gaussian Process hyperparmeters η, ρ^{-1}, and σ, and the negative binomial over-dispersion ϕ^{-1} parameter to be at least somewhat correlated and not strongly identified, corresponding to significant curvature in the model posterior. Hamiltonian Monte Carlo (HMC) is a particularly effective sampling strategy for posteriors with this property (Betancourt and Girolami, 2015). *Stan*'s *NUTS* sampler performs full joint Bayesian estimation of all parameters using HMC. The α_ρ and β_ρ hyperparameters of the ρ^{-1} hyperprior distribution are fixed as described in § 3.1, but otherwise all parameters are sampled using joint transitions to efficiently explore correlations between the parameters.

We typically fit 8 independent chains of length 2000 iterations (following an equal number of *NUTS* warmup samples) in parallel using *Stan* and observe a typical execution time of 1 min. However, in order to obtain the high resolution 2D posterior histograms used in this work, we run a larger simulation of 20 chains of 4000 samples each. To optimize the performance of the posterior simulation under the *Stan* modeling language, we use the Cholesky factor transformed implementation of the normal distribution to calculate the likelihood. We use the *cov_exp_quad* function in *Stan* to implement the squared exponential covariance function, and we rescale ρ^{-1} by $2^{-1/2}$ to accommodate the difference between this implementation and our definition in § 3. We use a non-centered parameterization (see e.g. Papaspiliopoulos et al. 2003) for the Gaussian process, modeling the latent parameter $\tilde{y}$ as standard normal and then transforming to a sampled value for y by rescaling by the covariance matrix.

The MCMC trace shown in Figure S1 illustrates the high independence of samples achieved after the *NUTS* algorithm warm-up period, and the low variance in sampling distributions between chains.

While we chose the negative binomial to permit overdispersion in the annualized mass shooting rate beyond counting noise, as Figure S1 shows, the data provides strong evidence for small values of ϕ^{-1}, consistent with Poisson noise.

We assess MCMC convergence quantitatively using the Gelman-Rubin convergence di-

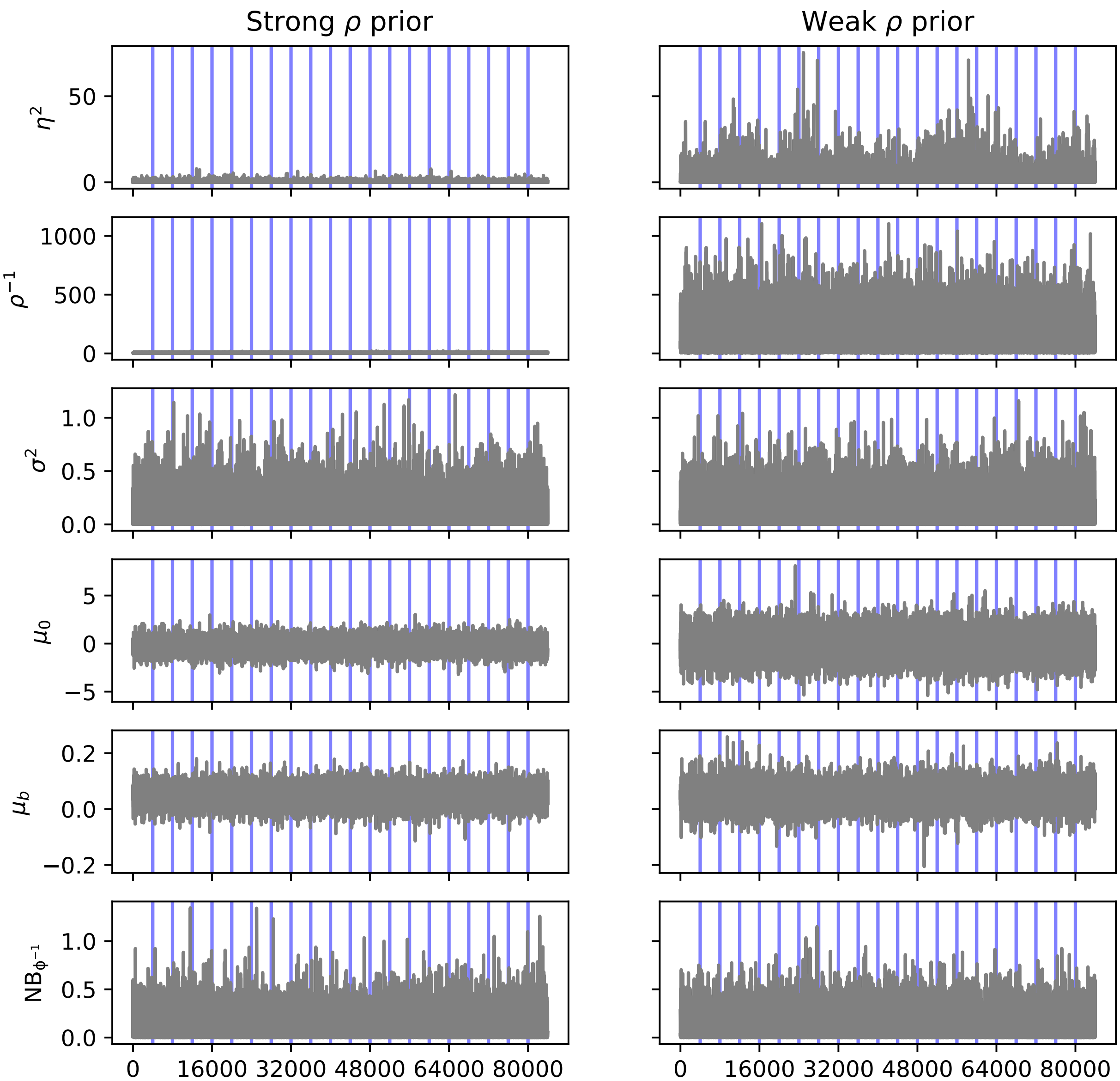

Figure S1: Traceplot showing the sequence of MCMC samples for the hyperparameters for the Gaussian process – η^2, which scales the strength of the covariance; ρ^{-1}, which scales the timescale; and σ^2, which sets the baseline level of variance–as well as the linear mean function–μ_0, the offset; μ_b, which sets the slope; and NB_ϕ^{-1}, which controls the overdispersion of the count data. The left panel shows the trace from the model with the strong prior on ρ^{-1} and the right side shows the model with the weak prior. The vertical lines divide individual chains from the MCMC simulation.

agnostic, $\hat{R}$, a comparison of within- to between-chain variance. As shown in Figure S2, we find that $\hat{R} \ll 1.05$ for all parameters, indicating a negligible discrepancy in the sampling distributions between chains. This holds for both models.

We perform a posterior predictive check by visualizing in Figure S3 the sampled values of z. The simulated data from this check realizes both a draw from the latent Gaussian process for the public mass shootings rate and the overdispersed counting noise of the negative binomial distribution. Visual inspection suggests that the observations simulated under the model show similar variation over time as the actual observations (first panel). We note that some realizations have annual counts at the later end of the modeled time range that exceed the largest observed annual count (7 public mass shootings). Some exceedence is expected given the counting noise, but this posterior predictive check could guide revision of the prior on the over-dispersion parameter or the choice of the negative binomial likelihood.

We note that the parameters of the linearized mean function are highly correlated, as shown in Figure S4. If the mean rate of public mass shootings at the beginning of the time series (μ_0) is inferred to be higher, then the increase in the mean function over time needed to explain the observations (μ_b) would be lower. However, at all probable values of μ_0, the distribution of μ_b is predominantly positive.

References

Betancourt, M. and Girolami, M. (2015). Hamiltonian monte carlo for hierarchical models. *Current trends in Bayesian methodology with applications*, 79:30.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, Articles*, 76(1):1–32.

Papaspiliopoulos, O., Roberts, Gareth, O., and Skold, M. (2003). Non-centered parameterisations for hierarchical models and data augmentation. In Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M., editors, *Bayesian Statistics 7*, pages 307–326. Oxford University Press, Oxford.

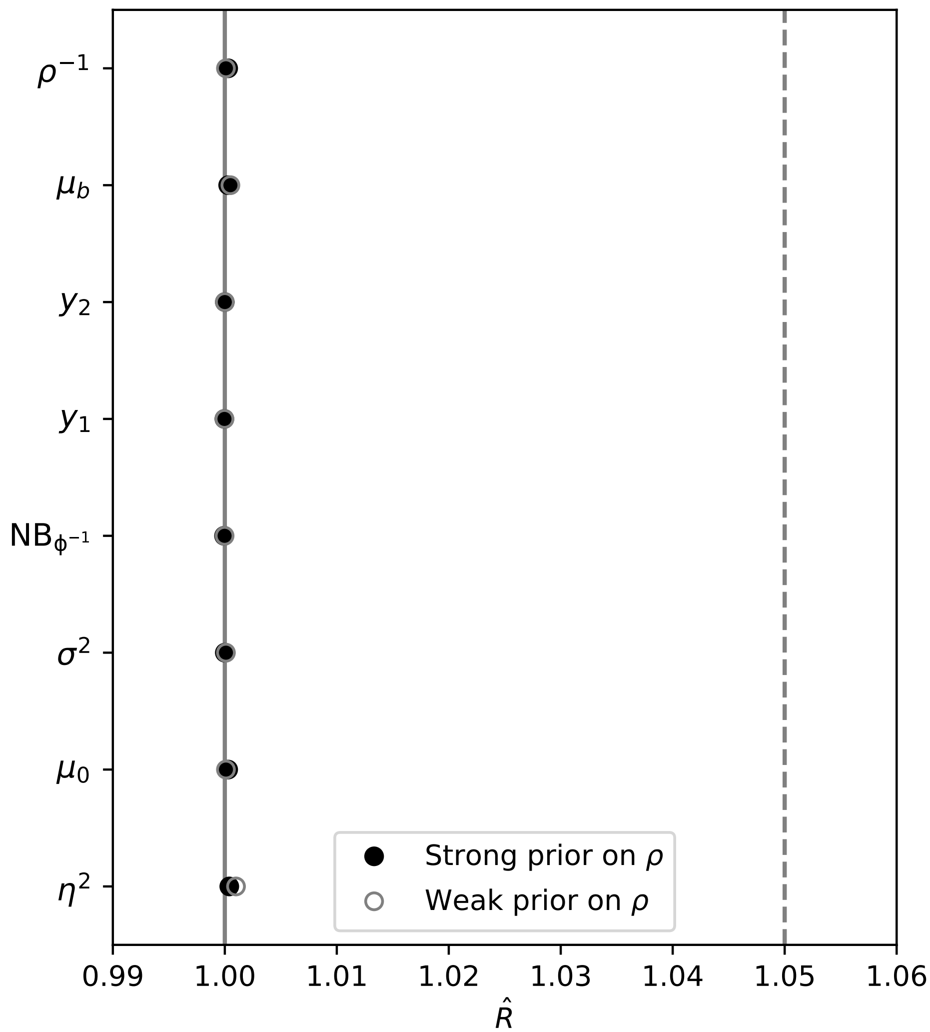

Figure S2: Comparison of the Gelman-Rubin convergence diagnostic, $\hat{R}$, across parameters for the models with strong (filled circles) and weak (open circles) priors on the timescale parameter ρ^{-1}. For the multi-valued latent parameters for the occurrence rate at the observed and interpolated points, y_1 and y_2 respectively, error bars are drawn to show the 90% interval of the distribution, though they are typically smaller than the marker size.

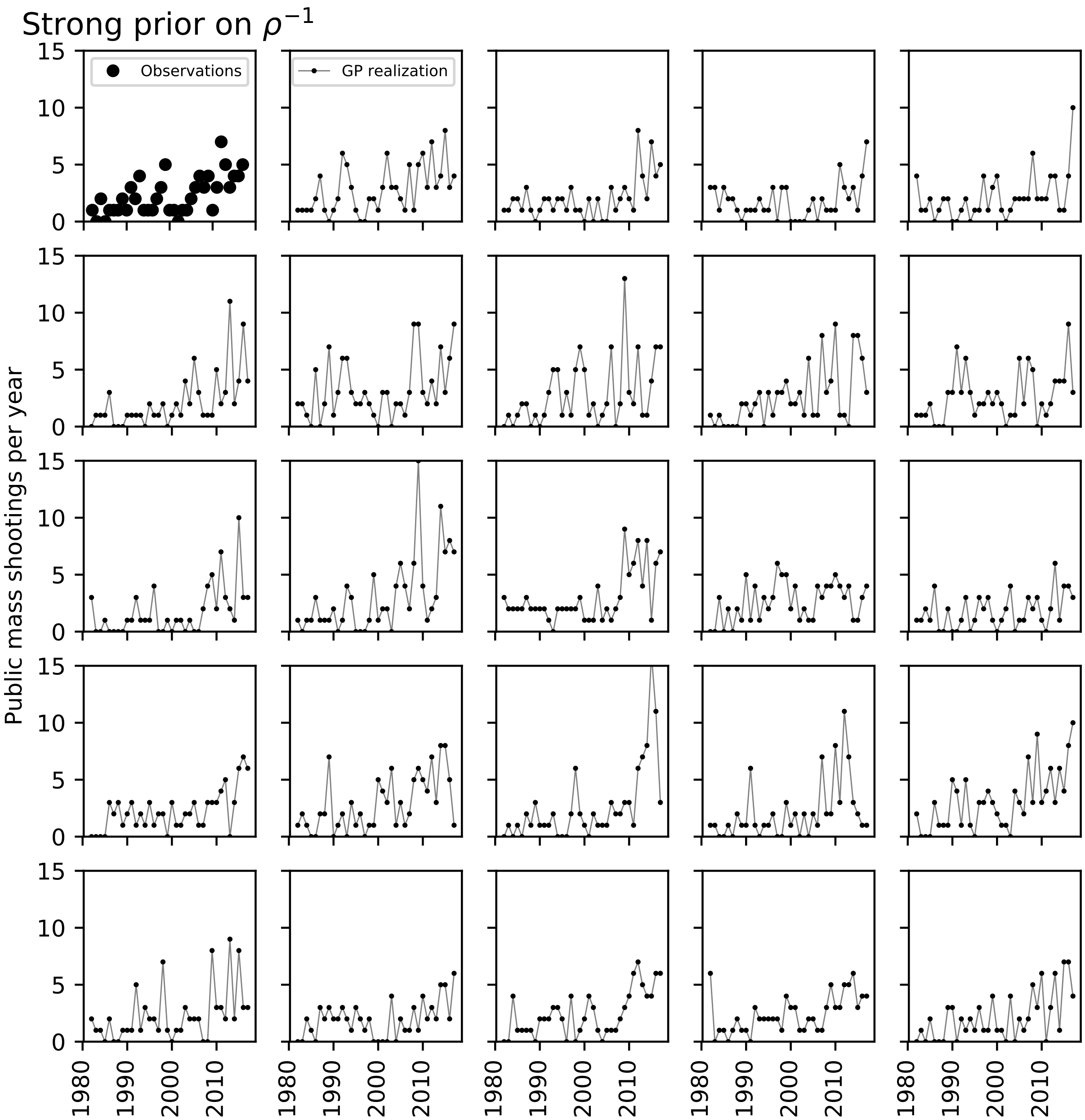

Figure S3: Posterior predictive check for the model with the strong prior on ρ^{-1}

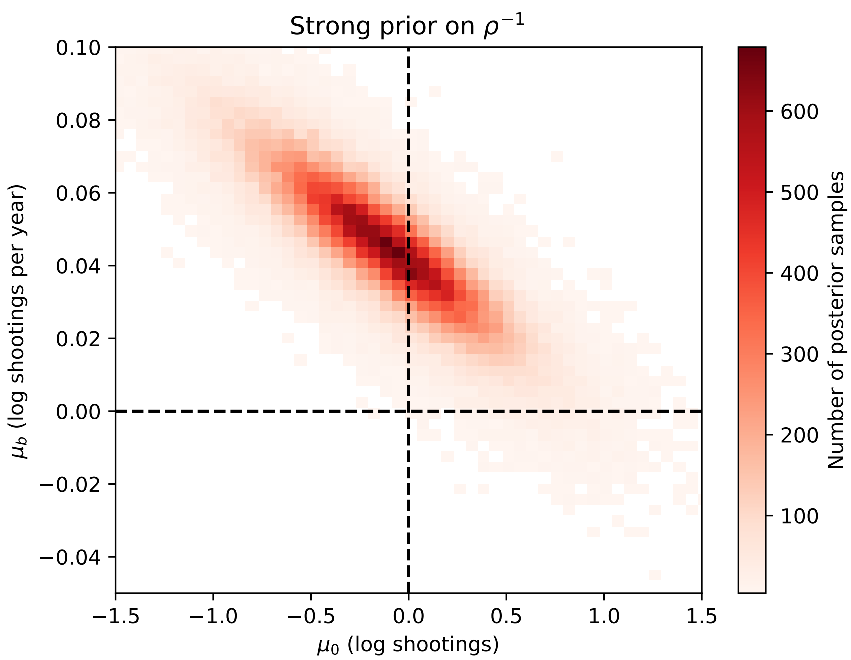

Figure S4: Two dimensional slice of the posterior distribution of the model with the strong prior on ρ^{-1}. The binned density of the posterior samples for the mean function intercept (μ_0) is compared to the mean function slope (μ_b) parameter.