STDGLM

```r
library(STDGLM)
```

Introduction

The `STDGLM` package provides a framework for fitting spatio-temporal dynamic generalized linear models. These models are useful for analyzing data that varies over both space and time, allowing for the incorporation of spatial and temporal dependencies in the modeling process. The package provides functions for fitting these models, as well as tools for visualizing and interpreting the results.

Installation

You can install the package from GitHub using the following command:

```r
if (!requireNamespace("devtools", quietly = TRUE)) {
  install.packages("devtools")
}
devtools::install_github("czaccard/STDGLM")
```

Detailed Explanation on supported STDGLMs

As for the current version of the package (`0.0.0.9000`), only Gaussian outcomes are supported. Specifically only dynamic linear models (DLMs) with the following specification can be handled:

$$y_t = x_t'\beta_t + z_t'\gamma + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2)$$

$$\beta_t = F_t\beta_{t-1} + \eta_t, \quad \eta_t \sim N(0, \Sigma_\eta)$$

where:

- y_t is the response variable at time t,
- x_t is a p-dimensional vector of covariates at time t (an intercept may or may not included here),
- β_t is the state vector at time t,
- z_t is a q-dimensional vector of covariates whose effects are constant (an intercept may or may not included here),
- γ is a vector of non-varying coefficients,
- ϵ_t is the observation error at time t,
- $F_t = \phi I_p$ is a transition matrix,
- Σ_η is the covariance matrix of the state evolution error η_t.

The state equation is adapted to account for spatial correlations in the state vector. Let p denote the number of *spatial units* (either georeferenced locations or areal units) where data is collected, then $\beta_t = (\beta_{1,t}, \dots, \beta_{p,t})'$, where $\beta_{i,t}$ is the effect of the covariate $x_{i,t}$ on the outcome at spatial unit $i = 1, \dots, p$ and time $t = 1, \dots, T$.

The state evolution covariance matrix Σ_η can be structured to reflect *spatial relationships*, e.g. by assuming an **exponential covariance function** if the data are point-referenced:

$$(\Sigma_\eta)_{ij} = \rho_1 \exp\left(-\frac{d_{ij}}{\rho_2}\right), \quad d_{ij} = \|s_i - s_j\|$$

where ρ_1 is the partial sill, ρ_2 is the range parameter, and d_{ij} is the Euclidean distance between locations s_i and s_j. At the moment, this is the only supported covariance structure for point-referenced data.

In this case, the evolution error η_t is assumed to be a zero-mean Gaussian process with exponential covariance matrix parameteterized by ρ_1 and ρ_2, which we will denote as $\eta_t \sim GP(0, \rho_1, \rho_2; exp)$.

If the data are areal, a **proper conditional autoregressive (PCAR) covariance structure** is assumed:

$$\Sigma_\eta = \rho_1 \left(D_w - \rho_2 W\right)^{-1}$$

where W is a binary adjacency matrix, D_w is a diagonal matrix with row sums of W on the diagonal, and ρ_1 and ρ_2 are the conditional variance and autocorrelation parameters, respectively.

In this case, the evolution error η_t follows a zero-mean PCAR process, and we will denote this as $\eta_t \sim PCAR(0, \rho_1, \rho_2)$.

ANOVA Decomposition of the State Vector

The function `stdglm` allows for the decomposition of the state vector into components that can be interpreted as contributions from different sources of variability. The state vector β_t is decomposed as follows:

$$\beta_{i,t} = \overline{\beta} + \beta_i^{(\mathsf{S})} + \beta_t^{(\mathsf{T})} + \beta_{i,t}^{(\mathsf{ST})}$$

where:

- $\overline{\beta}$ is the overall mean effect,
- $\beta_i^{(\mathsf{S})}$ is the spatial effect at location i,
- $\beta_t^{(\mathsf{T})}$ is the temporal effect at time t,
- $\beta_{i,t}^{(\mathsf{ST})}$ is the interaction effect between space and time at location i and time t.

Bayesian Hierarchical Structure

The Bayesian model is as follows, for $t = 1, \dots, T$:

$$
\begin{aligned}
y_t &\sim N(x_t'\beta_t + z_t'\gamma, \sigma^2_\epsilon) \\
\beta_t &= 1_p\overline{\beta} + \beta^{(\mathsf{S})} + 1_p\beta_t^{(\mathsf{T})} + \beta_t^{(\mathsf{ST})} \\
\overline{\beta} &\sim N(0, V_\gamma) \\
\beta^{(\mathsf{S})} &= (\beta_1^{(\mathsf{S})}, \dots, \beta_p^{(\mathsf{S})})' \sim GP(0, \rho_1^{(\mathsf{S})}, \rho_2^{(\mathsf{S})}; exp) \text{ or } PCAR(0, \rho_1^{(\mathsf{S})}, \rho_2^{(\mathsf{S})}) \\
\beta_t^{(\mathsf{T})} &\sim N(\phi^{(\mathsf{T})}\beta_{t-1}^{(\mathsf{T})}, V_\beta^{(\mathsf{T})}) \\
\beta_t^{(\mathsf{ST})} &= (\beta_{1,t}^{(\mathsf{ST})}, \dots, \beta_{p,t}^{(\mathsf{ST})})' \sim GP(\phi^{(\mathsf{ST})}\beta_{t-1}^{(\mathsf{ST})}, \rho_1^{(\mathsf{ST})}, \rho_2^{(\mathsf{ST})}; exp) \text{ or } PCAR(\phi^{(\mathsf{ST})}\beta_{t-1}^{(\mathsf{ST})}, \rho_1^{(\mathsf{ST})}, \rho_2^{(\mathsf{ST})}) \\
\gamma &\sim N(0, V_\gamma) \\
\sigma^2_\epsilon &\sim IG(a_\epsilon, b_\epsilon) \\
\rho_1^{(\mathsf{S})} &\sim IG(0.01, 0.01) \\
\rho_2^{(\mathsf{S})} &\sim U(a_\rho, b_\rho) \\
\rho_1^{(\mathsf{ST})} &\sim IG(0.01, 0.01) \\
\rho_2^{(\mathsf{ST})} &\sim U(a_\rho, b_\rho) \\
\phi^{(\mathsf{T})} &\sim TN_{(-1,1)}(0, 1) \\
\phi^{(\mathsf{ST})} &\sim TN_{(-1,1)}(0, 1) \\
V_\beta^{(\mathsf{T})} &\sim IG(a^{(\mathsf{T})}, b^{(\mathsf{T})})
\end{aligned}
$$

where $TN_{(q,r)}(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2 truncated to the interval (q, r). The hyperparameters a_ρ and b_ρ depend on the type of spatial data. If the data are point-referenced, they are set based on the minimum and maximum distances between points, respectively. If the data are areal, $a_\rho = 0.1$ and $b_\rho \to 1$.