A Bayesian Semiparametric Approach for Trend-Seasonal Interactions: an Application to Migration Forecasts. Additional Material

Alice Milivinti and Giacomo Benini

November 7, 2018

1   Bayesian Analysis

In this section we provide an overview of the Bayesian method on which our forecasts are based on. We illustrate how Bayesian techniques are implemented starting by rewriting equation (6) of the 1st order Fourier in matrix form

$$\mathbf{y}_t = \varphi \mathbf{x}_t + \epsilon_t \qquad \epsilon_t \sim N(0, \sigma_\epsilon^2), \tag{1}$$

where $\varphi = [\beta_0, \beta_1, \beta_{21}, \beta_{31}]^T$ and $\mathbf{x}_t = [1, \text{trend}_t, \cos_{1t} + \sin_{1t}, \text{trend}_t * (\cos_{1t} + \sin_{1t})]$.

STEP I Assume a prior probability distribution for $\varphi \sim P(\varphi|\alpha)$, where α are the parameters, which identify the prior distribution, i.e. the hyperparameter. The prior is the probability of φ which reflects the researcher's expectations.

STEP II Use the collected data $\mathbf{x}_t$ to compute the likelihood of the sample,

$$P(\mathbf{x_t}|\varphi) = \prod_{t=1}^{T} g(\mathbf{x}_t|\varphi), \tag{2}$$

where $g(\cdot)$ is the probability density function of the data.

STEP III According to the Bayes' rule, update the prior beliefs using the information contained in the likelihood obtaining the posterior distribution

$$P(\varphi|\mathbf{x_t}, \alpha) = \frac{P(\mathbf{x_t}|\varphi)P(\varphi|\alpha)}{P(\mathbf{x_t}|\alpha)} \tag{3}$$

that is, the probability of observing the hypothesis α given the observed data $\mathbf{x_t}$. Given the difficulty to accurately compute the marginal likelihood $P(\mathbf{x_t}|\alpha)$, the posterior is often approximated by $P(\varphi|\mathbf{x_t}, \alpha) \propto P(\mathbf{x_t}|\varphi)P(\varphi|\alpha)$.

The formalization of the Bayesian method shows how this methodology embodies the forecast principles combining probabilistic statements about the model's parameters φ and the available

data $(\mathbf{y}_t, \mathbf{x}_t)$ (Geweke & Whiteman, 2006). The resulting predictive probability is obtained by integrating over the posterior distribution,

$$G(\hat{\mathbf{y}}_\mathbf{t}|\mathbf{x}_\mathbf{t}, \alpha) = \int_\varphi P(\mathbf{y}_\mathbf{t}|\varphi)P(\varphi|\mathbf{x}_\mathbf{t}, \alpha)\mathrm{d}\varphi. \tag{4}$$

To compute such integrations, analytical solutions are generally hard to find and therefore numerical approximations are used. The procedure just described is able to produce accurate forecasts for parametric models. However, it cannot be implemented for the semiparametric specifications. The next section illustrates how to adapt these three steps to equations where at least one of the arguments is a nonparametric function.

1.1 Nonparametric Bayesian Techniques

Unlike equation (1), the semiparametric specifications require a prior distribution on a function. This reflects the need to make an assumption about every point in some continuous input space to which associate a certain statistical process. One of the most common choices is to assume that all the arguments are normally distributed random variables, so that the functions $f_1(\cdot), f_{21}(\cdot), f_{31}(\cdot)$ follow a Gaussian Process (GP),

$$f_1(\cdot) \sim GP(\cdot|0, \sigma^2_{f_1}) \quad f_{21}(\cdot) \sim GP(\cdot|0, \sigma^2_{f_{21}}) \quad f_{31}(\cdot) \sim GP(\cdot|0, \sigma^2_{f_{31}}). \tag{5}$$

The practical application of this relatively straightforward generalization of **STEP I** is rather convoluted and depends crucially upon the way the $f(\cdot)$s are computed. Therefore, before seeing how (5) allows to replicate the procedure illustrated in the previous section, it is necessary to decide how to estimate $\hat{f}_1(\cdot), \hat{f}_{21}(\cdot), \hat{f}_{31}(\cdot)$. Among the different options available to fit unknown functions, thin plate regression splines deliver the most efficient smoothers (Wood, 2003). Therefore, we use them to construct composite curves for which the last point of the first curve coincides with the first point of the second curve such that these two points have the same tangent (first derivative) and curvature (second derivative). Given the aim to re-adapt **STEP I** to the "prior assumption" (5), we re-write equation

$$y_t = \beta_0 + f_1(\mathrm{trend}_t) + f_{21}(\cos_{1t} + \sin_{1t}) + f_{31}(\mathrm{trend}_t, \cos_{1t} + \sin_{1t}) + \epsilon_t, \tag{6}$$

in matrix form,

$$\mathbf{y}_t = h(\mathbf{p}_t) + \epsilon_t. \tag{7}$$

Thin-plate regression splines can be estimated minimizing for $\mathbf{h}$

$$||\mathbf{y} - \mathbf{h}||^2 + \lambda J_{md}(h), \tag{8}$$

where $\mathbf{y}$ is a vector containing the data points $[y_1, y_2, ..., y_T]$, $\mathbf{h} = [f_1(\mathbf{p}_1), f_{21}(\mathbf{p}_2), f_{31}(\mathbf{p}_3)]$, with $\mathbf{p}_1 = [\mathrm{trend}_t]$, $\mathbf{p}_2 = [\cos_{1t} + \sin_{1t}]$ and $\mathbf{p}_3 = [\mathrm{trend}_t, \cos_{1t} + \sin_{1t}]$, λ is a smoothing parameter and $J_{md}(h)$ is a penalty function which measures the wiggliness of h. The particular form taken by

$J_{md}(h)$ is

$$J_{md} = \int \cdots \int_{\Re^d} \sum_{\nu_1+\ldots\nu_d=m} \frac{m!}{\nu_1!\ldots\nu_d!} \left(\frac{\partial^m h}{\partial p_1^{\nu_1}\ldots\partial p_d^{\nu_d}}\right)^2 dp_1\ldots dp_d, \tag{9}$$

where $T + 1$ is a sequence of non-decreasing real numbers which define the domain of each basis function (Duchon, 1977).

Minimization (8) includes the smoothing parameter $\lambda \in [0, \infty)$ which controls for the trade-off between model fit and smoothness. If $\lambda \to \infty$, $\hat{h}$ is a straight line estimate converging to the linear least squares estimate. If $\lambda \to 0$, $\hat{h}$ is an un-penalized regression spline estimate suffering from an over-fitting issue (De Boor, 1978). Solving (8), for a fixed λ, allows to write the estimate of h as a thin plate regression splne,

$$\hat{h}_\lambda(\mathbf{p}) = \sum_{j=1}^{M} \theta_j \phi_j(\mathbf{p}) + \sum_{t}^{T} \delta_t \eta_{md}(\|\mathbf{p} - \mathbf{p}_t\|), \tag{10}$$

where δ_t and θ_j are coefficients to be estimated and $\phi_j(\mathbf{p})$ is the vector of $M = \binom{m+d-1}{d}$ unpenalized polynomials of $\mathbf{p}$ with degrees up to $m-1$. Beside $\phi_j(\mathbf{p})$, $\eta_{m,d}$, which is a function of the Euclidean distance r between any $\mathbf{p}$ and the observed $\mathbf{p}_t$ values, models the additional non-linearity:

$$\eta_{m,d}(r) = \begin{cases} \frac{(-1)^{m+1+d/2}}{2^{2m-1}\pi^{d/2}(m-1)!(m-d/2)!} r^{2m-d} \log(r) & \text{if } d \text{ even} \\ \frac{\Gamma(d/2-m)}{2^{2m}\pi^{d/2}(m-1)!} r^{2m-d} & \text{if } d \text{ odd.} \end{cases} \tag{11}$$

By defining the penalty matrix $\mathbf{E}$ such that each entry $e_{tj} \equiv \eta_{m,d}(\|\mathbf{p}_t - \mathbf{p}_j\|)$, the function estimate $\hat{h}_\lambda$ is then obtained by minimizing equation (12) with respect to δ and θ :

$$\min \|\mathbf{y} - \mathbf{C}\theta - \mathbf{E}\delta\|^2 + \lambda\delta'\mathbf{E}\delta, \quad \mathbf{C}'\delta = \mathbf{0}, \tag{12}$$

where, $\mathbf{y}$ is the responses' vector, $\mathbf{C}$ is the matrix whose rows are $\phi_j(\mathbf{p})$, θ and δ are vectors of coefficients θ_j and δ_t (Wood, 2006). Besides the great flexibility qualities of the thin plate regression spline as a smoother, the estimate of $\hat{h}_\lambda$ engages as many parameters as the number of data points, generating important computational costs. Therefore, solving (θ, δ) with an optimum λ^* becomes difficult for large problems. Being $\mathbf{E}$ a symmetric and non-negative definite matrix, the eigen-decomposition can be written as $\mathbf{E} = \mathbf{UDU}'$, where $\mathbf{D}$ is the diagonal matrix of d_t eigenvalues of $\mathbf{E}$ and $\mathbf{U}$ is the corresponding eigenvectors matrix. The truncated eigen-decomposition produces $\tilde{\mathbf{E}}_k$, which is an approximation to $\mathbf{E}$ such that

$$\tilde{\mathbf{E}}_k = \mathbf{U}_k \mathbf{D}_k \mathbf{U}'_k$$

where $\mathbf{D}_k$ is a diagonal matrix that contains the k most extreme eigenvalues in descending order of absolute values (i.e. the top right $k \times k$ sub-matrix of $\mathbf{D}$): $|\tilde{d}_1| > \cdots > |\tilde{d}_k|$. $\mathbf{U}_k$ is the eigenvectors matrix corresponding to the eigenvalues in $\mathbf{D}_k$. The approximation $\tilde{\mathbf{E}}_k$ brings some advantages. First it reduces the dimension from $T \times T$ of $\mathbf{E}$ to $T \times k$. Second, $\tilde{\mathbf{E}}_k$ minimizes the spectral norm $\|\mathbf{E} - \mathbf{F}_k\|_2$ between $\mathbf{E}$ and all rank k matrices $\mathbf{F}_k$. Third, $\tilde{\mathbf{E}}_k$ minimizes the worst possible change in the shape of the spline introduced by the eigen-space truncation measured by

$$\max_{\delta \neq 0} \frac{\delta'(\mathbf{E} - \mathbf{G}_k)\delta}{\|\delta\|^2},$$

where $\mathbf{G}_k$ is formed by any k eigenvalues and corresponding eigenvectors (Wood, 2003). Given $\mathbf{E} \approx \tilde{\mathbf{E}}_k$ and $\tilde{\mathbf{E}}_k = \mathbf{U}_k \mathbf{D}_k \mathbf{U}_k'$ and letting $\delta_k = \mathbf{U}_k' \delta$, the minimization problem becomes

$$\min \|\mathbf{y} - \mathbf{C}\theta - \mathbf{U}_k \mathbf{D}_k \delta_k\|^2 + \lambda \delta_k' \mathbf{D}_k \delta_k \quad \text{subject to} \quad \mathbf{C}' \mathbf{U}_k \delta_k = 0.$$

The constrained optimization problem can be converted into an unconstrained one by using any orthogonal column basis $\mathbf{Z}$. One way to form $\mathbf{Z}$ is via the QR decomposition of $\mathbf{U}_k' \mathbf{C}$:

$$\mathbf{U}_k' \mathbf{C} = [\mathbf{Q_1} \mathbf{Q_2}] \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix}.$$

Letting $\mathbf{Z} = \mathbf{Q_2}$, it is verified that

$$\mathbf{T}' \mathbf{U}_k \mathbf{Z} = \mathbf{R}' \mathbf{Q_1}' \mathbf{Q_2} = \mathbf{0}.$$

Therefore, it is true that $\delta_k = \mathbf{Z}\tilde{\delta}$ for δ_k such that $\mathbf{C}' \mathbf{U}_k \delta_k = 0$. Now the problem becomes the unconstrained optimization,

$$\min \|\mathbf{y} - \mathbf{C}\theta - \mathbf{U}_k \mathbf{D}_k \mathbf{Z}\tilde{\delta}\|^2 + \lambda \tilde{\delta}' \mathbf{Z}' \mathbf{D}_k \mathbf{Z}\tilde{\delta}.$$

By defining

$$\kappa = \begin{bmatrix} \theta \\ \tilde{\delta} \end{bmatrix}, \quad \mathbf{p} = [\mathbf{C} : \mathbf{U}_k \mathbf{D}_k \mathbf{Z}], \quad \text{and} \quad \mathbf{S} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}' \mathbf{D}_k \mathbf{Z} \end{bmatrix}$$

the optimization can be simplified as $\min \|\mathbf{y} - \mathbf{p}\kappa\|^2 + \lambda \kappa' \mathbf{S}\kappa$ with respect to κ. This last reformulation of the minimization problem allows to rethink (5) in terms of assumptions about the distribution of λ and of κ. In particular, if

$$f_1(\cdot) \sim GP(\cdot|0,1) \quad f_{21}(\cdot) \sim GP(\cdot|0,1) \quad f_{31}(\cdot) \sim GP(\cdot|0,1), \tag{13}$$

then it is possible to assume as priors $\kappa \sim N(0,1)$ and $\lambda \sim N(0,\sigma_\lambda^2)$ and start the estimation process.

2 Simulations Results

This section provides the simulations' codes as well as the results' visualization of the fitting and forecasts of the linear, the semi- and the non- parametric models with and without interaction (see section 2.2 of the paper).

```
freq <- 12
years <- 30
N <- freq*years
x <- (1:N)/freq
```

```r
set.seed(1)


##########################################################################
# SIMULATION 1.1: 1st order Fourier trend, seasonality no interaction
##########################################################################


# Trend
trend_11 <- cumsum(cumsum(rnorm(N))+rnorm(N))

# Seasonality
sea_11 <- .3 + 3*cos(2*pi*x) + 6*sin(2*pi*x)

# Time Series data
sim_11 <- as.data.frame(abs(0.4*trend_11 + 10*sea_11 + .2*rnorm(N)))


##########################################################################
# SIMULATION 2.1: 2nd order Fourier trend, seasonality no interaction
##########################################################################
set.seed(1)
# Trend
trend_21 <- cumsum(cumsum(rnorm(N))+rnorm(N))

# Seasonality
sea_21 <- .3 + .5*cos(2*pi*x) + 1.2*sin(2*pi*x) +
  .6*cos(2*2*pi*x) + .2*sin(2*2*pi*x)

# Time Series data
sim_21 <- as.data.frame(abs(0.1*trend_21 + 40*sea_21 + .2*rnorm(N)))



##########################################################################
# SIMULATION 1.2: 1st order Fourier trend, seasonality with interaction
##########################################################################
set.seed(1)

# Trend
trend_12 <- cumsum(cumsum(rnorm(N))+rnorm(N))

# Seasonality
sea_12 <- .3 + 3*cos(2*pi*x) + 6*sin(2*pi*x)

# Time Series data
```

```r
sim_12 <- as.data.frame(0.4*trend_12 + 24*sea_12 + 0.8*trend_12*sea_12 + .2*rnorm(N))


#########################################################################
# SIMULATION 2.2: 2nd order Fourier trend, seasonality with interaction
#########################################################################

set.seed(1)
# Trend
trend_22 <- cumsum(cumsum(rnorm(N))+rnorm(N))
# Seasonality
sea_22 <- .3 + .5*cos(2*pi*x) + 1.2*sin(2*pi*x) +
  .6*cos(2*2*pi*x) + .2*sin(2*2*pi*x)

# Time Series data
sim_22 <- as.data.frame(0.4*trend_22 + 24*sea_22 + 0.8*trend_22*sea_22 + .2*rnorm(N))

#############################################################################
# SIMULATION 1.3: 1st order Fourier trend, seasonality with complex interaction
#############################################################################
set.seed(1)

# Trend
trend_13 <- sqrt(cumsum(cumsum(rnorm(N)))+rnorm(N))

# Seasonality
sea_13 <- .3 + 3*cos(2*pi*x) + 6*sin(2*pi*x)

# Time Series data
sim_13 <- as.data.frame(3 + 15*trend_13 + 10*exp(sea_13)*trend_13 + .4*rnorm(N))



#############################################################################
# SIMULATION 2.3: 2nd order Fourier trend, seasonality with complex interaction
#############################################################################
set.seed(1)
# Trend
trend_23 <- sqrt(cumsum(cumsum(rnorm(N)))+rnorm(N))

# Seasonality
sea_23 <- .3 + .5*cos(2*pi*x) + 1.2*sin(2*pi*x) +
  .6*cos(2*2*pi*x) + .2*sin(2*2*pi*x)
```

```
# Time Series data
sim_23 <- as.data.frame(3 + 15*trend_23 + 10*exp(sea_23)*trend_23 + .4*rnorm(N))
```
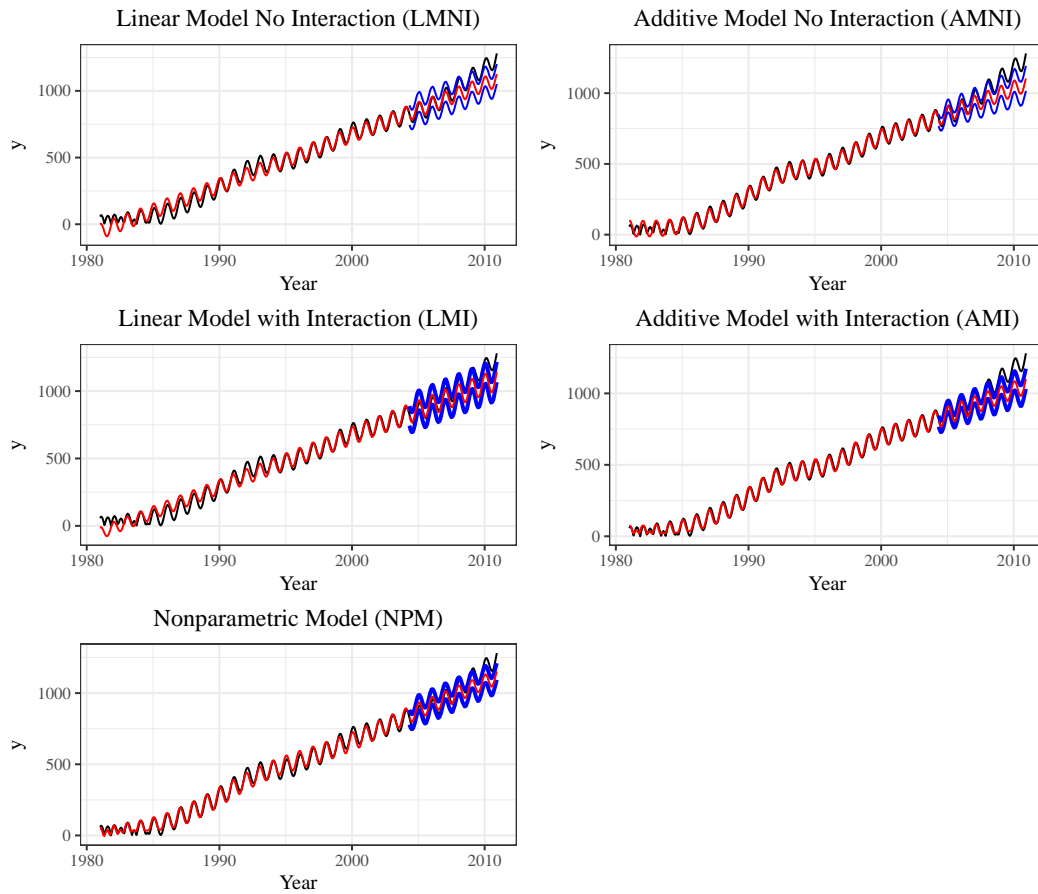
Figure 1: 1st order Fourier series with no interaction: fitting and forecasts results for different models.
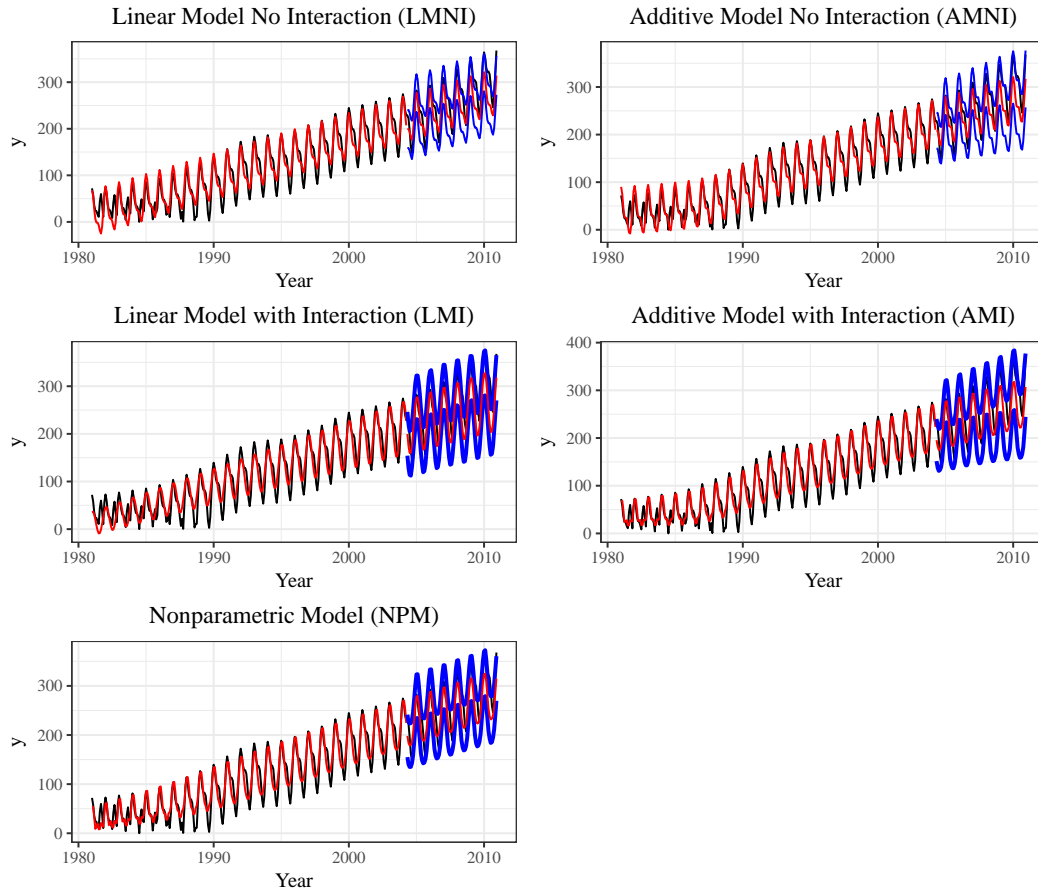
Figure 2: 2nd order Fourier series with no interaction: fitting and forecasts results for different models.
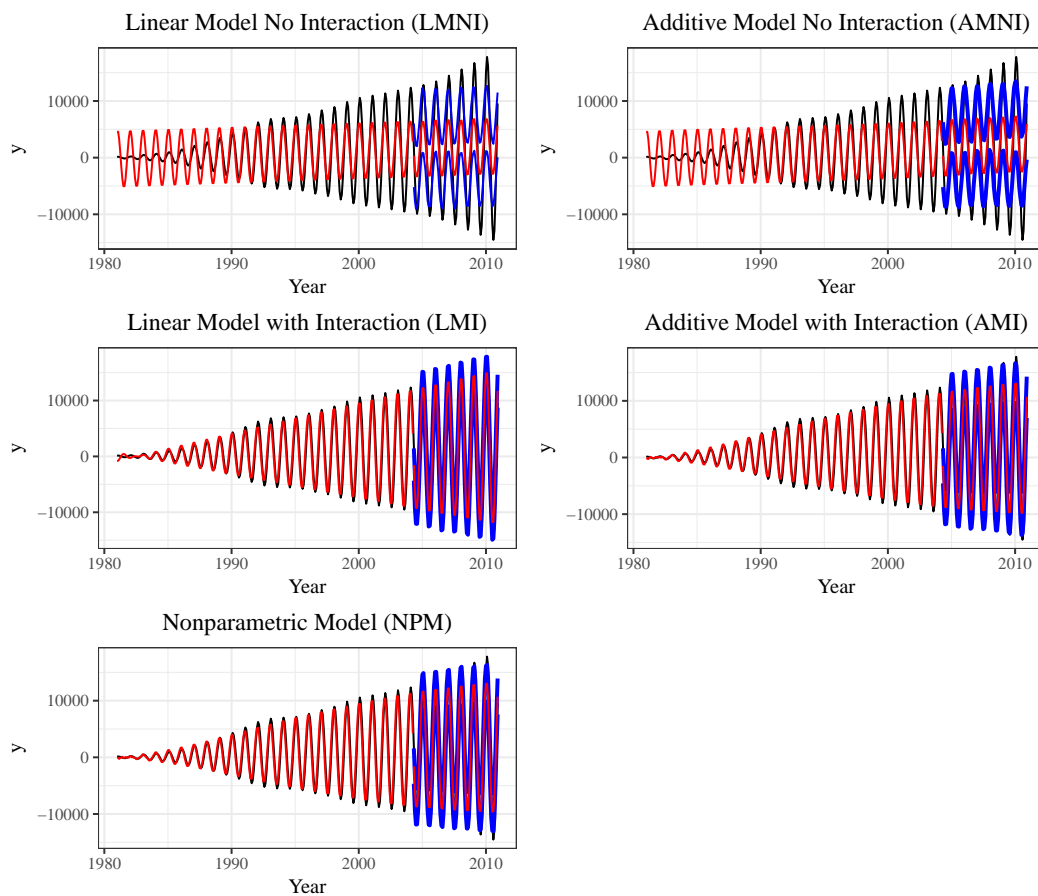
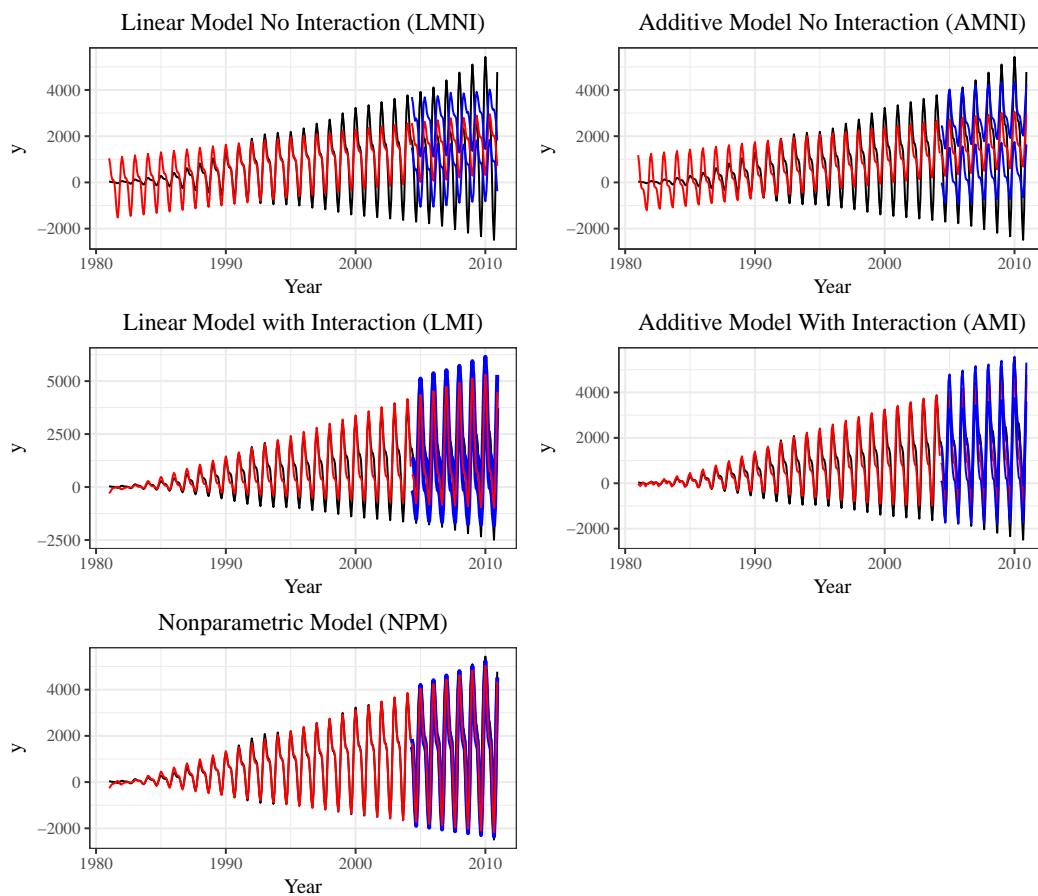Figure 3: 1st order Fourier series with interaction: fitting and forecasts results for different models.

Figure 4: 2nd order Fourier series with interaction: fitting and forecasts results for different models.
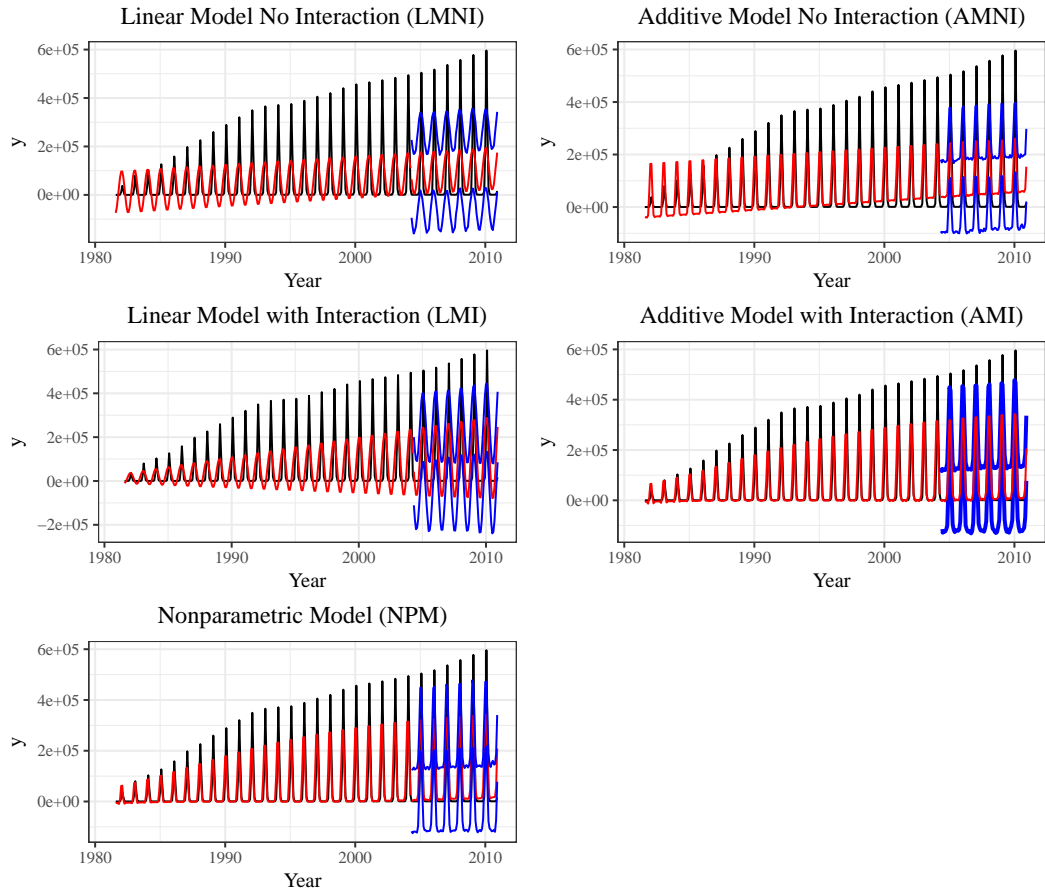
Figure 5: 1st order Fourier series with non-linear interaction: fitting and forecasts results for different models.
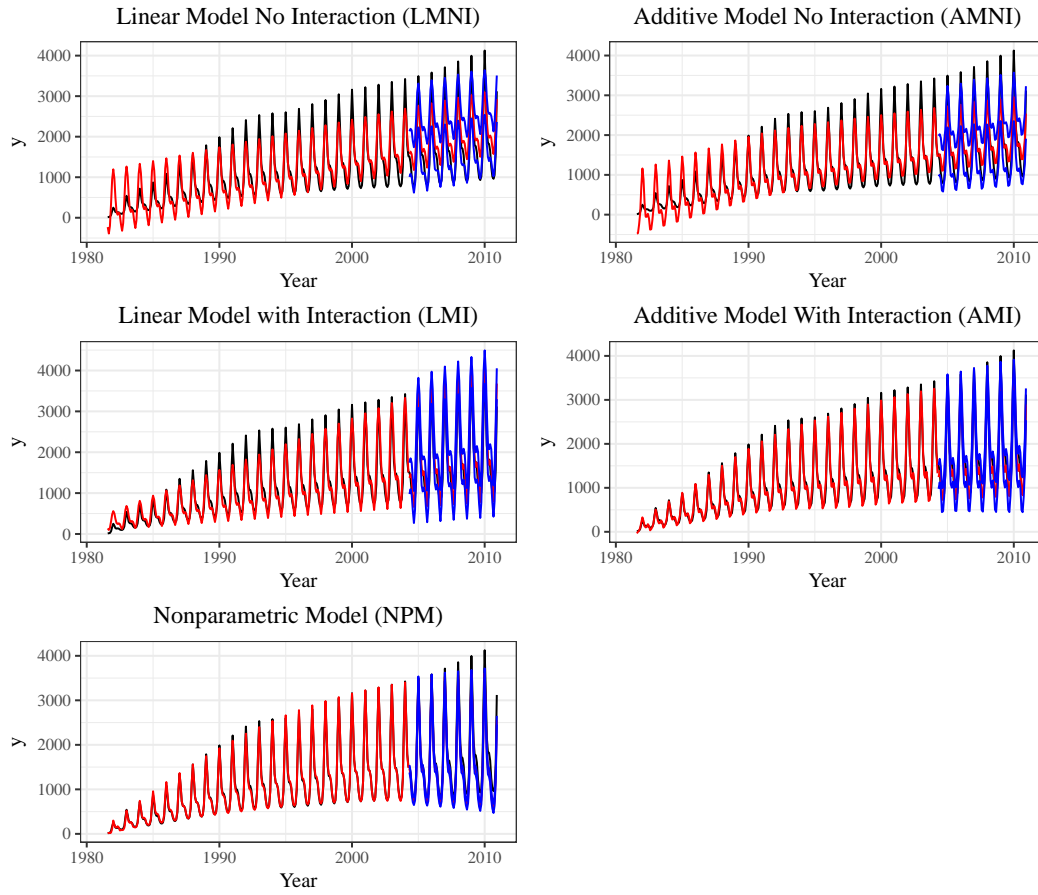
Figure 6: 2nd order Fourier series with non-linear interaction: fitting and forecasts results for different models.

3   Regression Results

This section reports the results of the Bayesian regressions fitted for the period 1981-2003 both for the time series case (Tables 1, 2, 3, 4 and 5), as well as for the longitudinal case (Tables 6, 7, 8, 9 and 10). For the time series case the models include autoregressive errors of order one (AR(1)). The estimates are obtained using 4 Markov chains, each with 4000 iterations, 2000 for warm-up and 2000 for validation, resulting in a total post warm-up sample of 8000 data points. The tables display, from the left to the right column, the potential scale reduction factor ($\hat{R}$), the mean (mean), the standard deviation (sd), the lower 2.5% quantile, the 50% quantile and the upper 97.5% quantile for all the models. A value of $\hat{R}$ equal to 1 indicates that each chain has stabilized and they are likely to have reached the target distribution[1].

In order to fully characterize the two semiparametric models, we add to the parametric values of the splines' coefficients a portray of the smooth terms, see Figure (7).

---

[1]A $\hat{R}$ significantly bigger than 1 indicates that the between-chain variance is substantially greater than the within-chain variance, warning for the need of a longer simulation.

Table 1: Linear Model without Interaction (LMNI) with AR(1) errors.

Parameter	$\hat{R}$	mean	sd	2.5%	50%	97.5%
Intercept	1.0	8.7	0.1	8.6	8.7	8.9
trend	1.0	0.4	0.2	0.0	0.4	0.7
$\sin_1 + \cos_1$	1.0	0.0	0.0	-0.0	0.0	0.1
$\sin_2 + \cos_2$	1.0	-0.0	0.0	-0.1	-0.0	0.0
α	1.0	0.5	0.1	0.4	0.5	0.6
σ_ϵ	1.0	0.3	0.0	0.2	0.3	0.3

Table 2: Additive Model without Interaction (AMNI) with AR(1) errors.

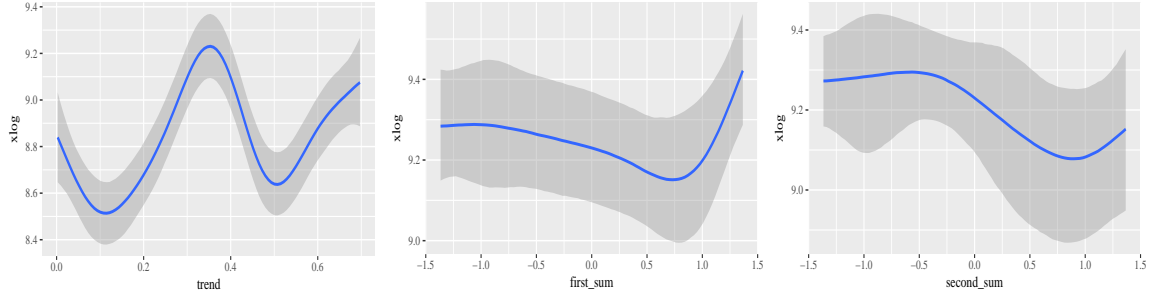| Parameter	$\hat{R}$	mean	sd	2.5%	50%	97.5%
Intercept	1.0	8.8	0.0	8.8	8.8	8.9
α	1.0	0.2	0.1	0.1	0.2	0.4
$\sigma_{\lambda_{\text{trend}}}$	1.0	2.7	0.8	1.5	2.5	4.7
$\sigma_{\lambda_{\sin_1 + \cos_1}}$	1.0	1.4	0.7	0.5	1.2	3.2
$\sigma_{\lambda_{\sin_2 + \cos_2}}$	1.0	1.0	0.7	0.3	0.8	2.8
σ_ϵ	1.0	0.2	0.0	0.2	0.2	0.2

Figure 7: Visualization of the non-linear functions of the additive model without interaction (AMNI). The left plot shows the monthly trend, the center plot $(\sin_1 + \cos_1)$ and the right plot $(\sin_2 + \cos_2)$. The central blue lines are the mean values of the dependent variable immigration (xlog) as a function of the x axis (trend, first sum=$(\sin_1 + \cos_1)$, second sum=$(\sin_2 + \cos_2)$. The dark grey bands depict the standard errors of the estimates and represent the credible intervals.

Table 3: Linear Model with Interaction (LMI) with AR(1) errors.

Parameter	$\hat{R}$	mean	sd	2.5%	50%	97.5%
Intercept	1.0	8.7	0.1	8.6	8.7	8.8
trend	1.0	0.4	0.2	0.1	0.4	0.7
$\sin_1 + \cos_1$	1.0	0.1	0.0	-0.0	0.1	0.2
$\sin_2 + \cos_2$	1.0	-0.0	0.0	-0.1	-0.0	0.1
$(\sin_1 + \cos_1)$*trend	1.0	-0.2	0.1	-0.5	-0.2	0.0
$(\sin_2 + \cos_2)$*trend	1.0	-0.1	0.1	-0.2	-0.1	0.1
α	1.0	0.5	0.1	0.4	0.5	0.6
σ_ϵ	1.0	0.3	0.0	0.2	0.3	0.3

Table 4: Additive Model with Interaction (AMI) with AR(1) errors.

Parameter	$\hat{R}$	mean	sd	2.5%	50%	97.5%
Intercept	1.0	8.8	0.0	8.8	8.8	8.9
α	1.0	0.2	0.1	0.1	0.2	0.4
$\sigma\lambda_{\text{trend}}$	1.0	2.7	0.8	1.5	2.5	4.6
$\sigma\lambda_{\sin_1 + \cos_1}$	1.0	0.9	0.8	0.0	0.8	2.8
$\sigma\lambda_{\sin_2 + \cos_2}$	1.0	0.8	0.7	0.0	0.6	2.5
$\sigma\lambda_{\text{trend} \sin_1 + \cos_1}$	1.0	0.2	0.1	0.0	0.2	0.5
$\sigma\lambda_{\text{trend} \sin_2 + \cos_2}$	1.0	0.1	0.1	0.0	0.1	0.4
σ_ϵ	1.0	0.2	0.0	0.2	0.2	0.2

Table 5: Nonparametric Model (NPM) with AR(1) errors.

Parameter	$\hat{R}$	mean	sd	2.5%	50%	97.5%
Intercept	1.0	8.9	0.0	8.8	8.9	8.9
α	1.0	0.6	0.0	0.5	0.6	0.7
$\sigma_{\lambda_{\text{trend},\sin_1+\cos_1,\sin_2+\cos_2}}$	1.0	0.1	0.0	0.1	0.1	0.2
σ_ϵ	1.0	0.2	0.0	0.2	0.2	0.2

Table 6: Linear Model without Interaction (LMNI) by Age.

Parameter	$\hat{R}$	mean	sd	2.5%	50%	97.5%
Intercept	1.0	5.8	0.0	5.7	5.8	5.8
trend	1.0	0.3	0.0	0.3	0.3	0.4
$\sin_1+\cos_1$	1.0	0.0	0.0	-0.0	0.0	0.0
$\sin_2+\cos_2$	1.0	-0.0	0.0	-0.0	-0.0	0.0
Age	1.0	-0.1	0.0	-0.1	-0.1	-0.1
σ_ϵ	1.0	0.9	0.0	0.9	0.9	0.9

Table 7: Additive Model without Interaction (AMNI) by Age.

Parameter	$\hat{R}$	mean	sd	2.5%	50%	97.5%
Intercept	1.0	3.4	0.0	3.4	3.4	3.4
$\sigma_{\lambda_{\text{trend}}}$	1.0	3.2	0.9	2.0	3.0	5.5
$\sigma_{\lambda_{\sin_1+\cos_1}}$	1.0	0.1	0.1	0.0	0.0	0.2
$\sigma_{\lambda_{\sin_2+\cos_2}}$	1.0	0.1	0.1	0.0	0.0	0.2
σ_ϵ	1.0	0.5	0.0	0.5	0.5	0.5
$\sigma_{\lambda_{Age}}$	1.0	4.4	2.0	2.2	3.9	9.5

Table 8: Linear Model with Interaction (LMI) by Age.

Parameter	$\hat{R}$	mean	sd	2.5%	50%	97.5%
Intercept	1.0	5.8	0.0	5.7	5.8	5.8
trend	1.0	0.3	0.0	0.3	0.3	0.4
$\sin_1+\cos_1$	1.0	-0.0	0.0	-0.0	-0.0	0.0
$\sin_2+\cos_2$	1.0	-0.0	0.0	-0.0	-0.0	0.0
$(\sin_1+\cos_1)$*trend	1.0	0.0	0.0	-0.0	0.0	0.1
$(\sin_2+\cos_2)$*trend	1.0	0.0	0.0	-0.0	0.0	0.1
Age	1.0	-0.1	0.0	-0.1	-0.1	-0.1
σ_ϵ	1.0	0.9	0.0	0.9	0.9	0.9

Table 9: Additive Model with Interaction (AMI) by Age.

Parameter	$\hat{R}$	mean	sd	2.5%	50%	97.5%
Intercept	1.0	3.4	0.0	3.4	3.4	3.4
$\sigma\lambda_{\text{trend}}$	1.0	3.5	1.0	2.1	3.3	5.8
$\sigma\lambda_{\sin_1 + \cos_1}$	1.0	0.1	0.1	0.0	0.1	0.3
$\sigma\lambda_{\sin_2 + \cos_2}$	1.1	0.1	0.1	0.0	0.0	0.5
$\sigma\lambda_{\text{trend}\sin_1 + \cos_1}$	1.0	0.0	0.0	0.0	0.0	0.0
$\sigma\lambda_{\text{trend}\sin_2 + \cos_2}$	1.0	0.0	0.0	0.0	0.0	0.0
$\sigma\lambda_{Age}$	1.0	5.3	1.3	3.5	5.0	8.5
σ_ϵ	1.0	0.5	0.0	0.5	0.5	0.5

Table 10: Nonparametric Model (NPM) by Age.

Parameter	$\hat{R}$	mean	sd	2.5%	50%	97.5%
Intercept	1.0	3.4	0.0	3.4	3.4	3.4
$\sigma\lambda_{\text{trend},\sin_1 + \cos_1,\sin_2 + \cos_2,Age}$	1.0	10.2	0.7	8.9	10.2	11.7
σ_ϵ	1.0	0.5	0.0	0.4	0.5	0.5

4 Diagnostics

The validation of a model is centered around the analysis of its residuals, i.e. the distance between observed and the fitted values. For each regression, we show four diagnostic plots. The first is a comparison between the standardized residuals and the fitted values (top-left). The second is a histogram of the standardized residuals (top-right). The third is a probability-probability plot (PP) which checks how close are the cumulative distribution functions obtained using the data set of the model (bottom-left). The last graph is a comparison of the fitted and the observed values (bottom-right).

All the models show standardized residuals which are rather random and do not form particular clusters. This regular behaviour is replicated by the histograms of the residuals which are all centered around zero and rather regular in their decline to extreme values. The PP plots show a cumulative distribution function of the model very close to the one of the data generating process (i.e. the fitted values are on the 45 degree line). The same is true for response versus fitted values, which are almost on the bissectrice of the first quadrant.

The results obtained using the disaggregate models are similar to the ones achieved using the aggregate specifications.
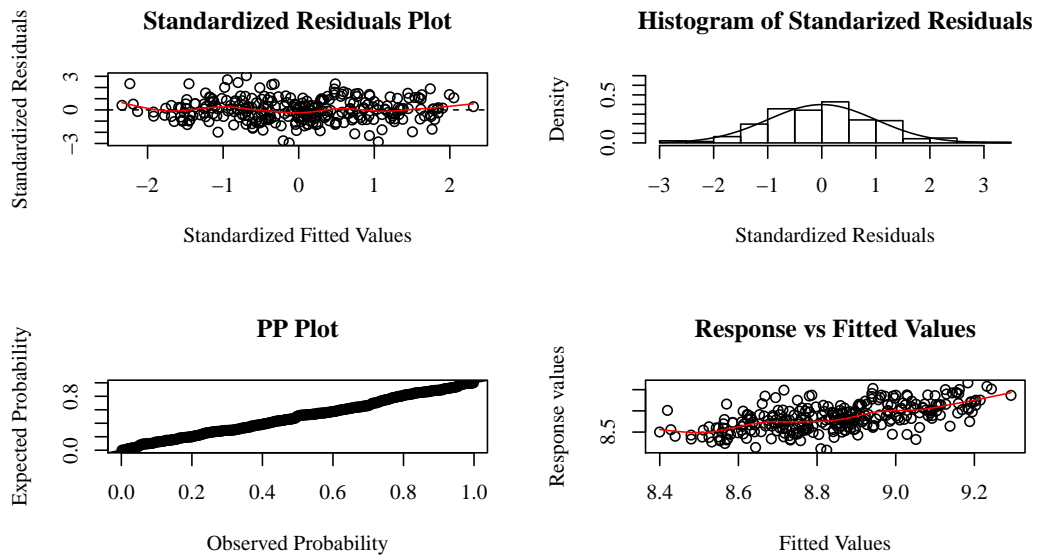
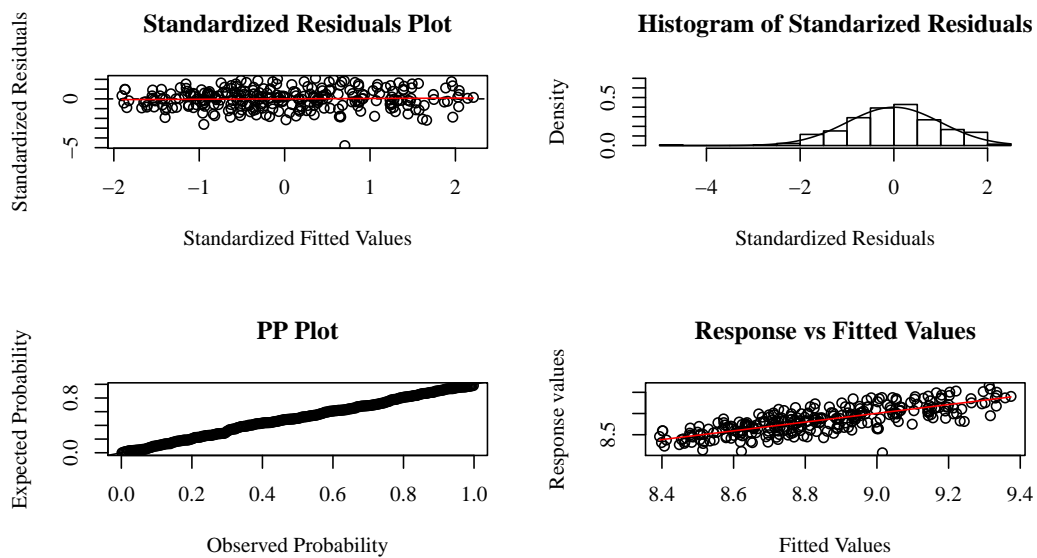Figure 8: Diagnostic Plot Linear Model without Interaction (LMNI).

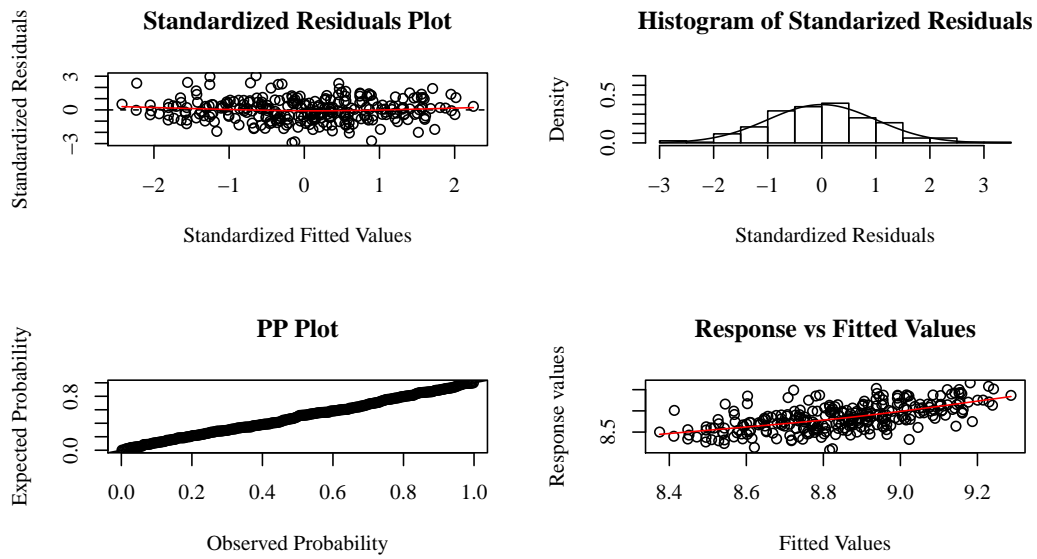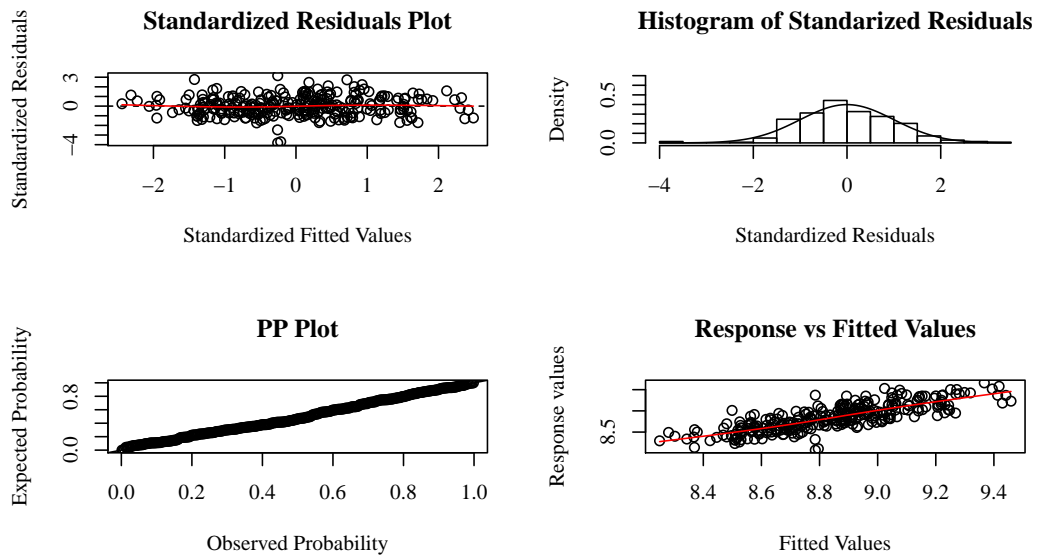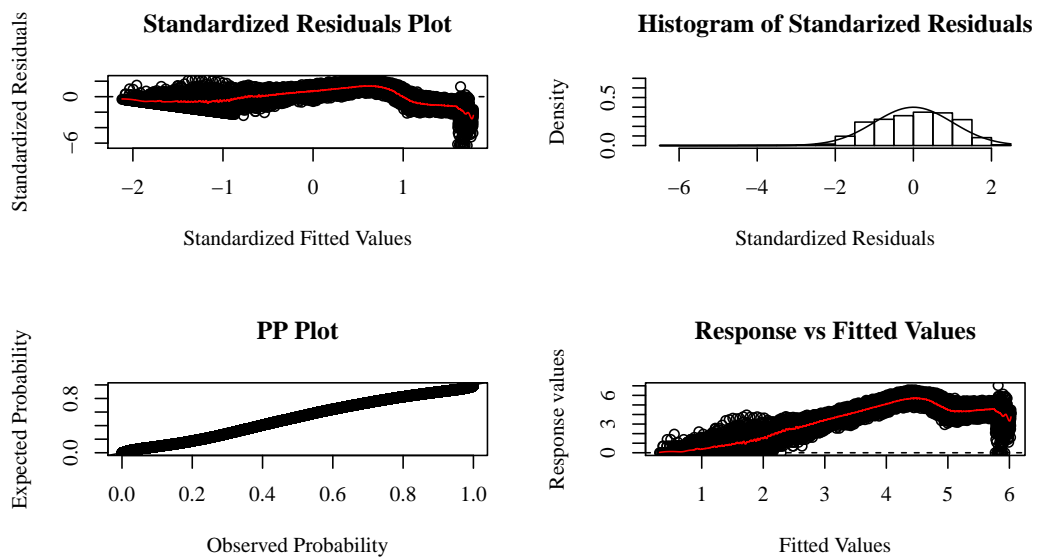Figure 9: Diagnostic Plot Additive Model without Interaction (AMNI).

Standardized Residuals Plot

Histogram of Standarized Residuals

PP Plot

Response vs Fitted Values

Figure 10: Diagnostic Plot Linear Model with Interaction (LMI).

Standardized Residuals Plot

Histogram of Standarized Residuals

PP Plot

Response vs Fitted Values
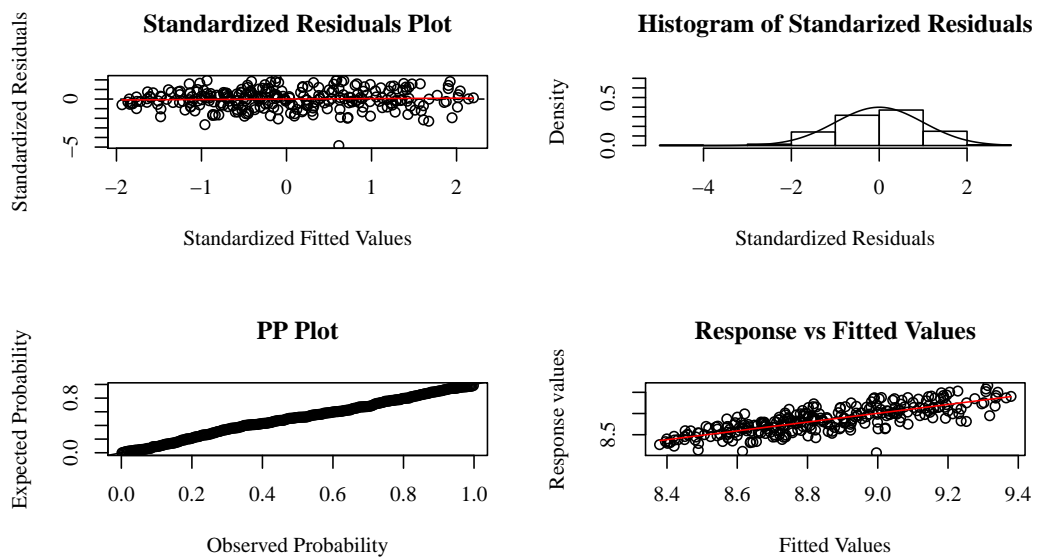
Figure 11: Diagnostic Plot Additive Model with Interaction (AMI).

18

Standardized Residuals Plot

Histogram of Standardized Residuals

PP Plot

Response vs Fitted Values

Figure 12: Diagnostic Plot Nonparametric Model with Interaction (NPM).

Standardized Residuals Plot

Histogram of Standardized Residuals
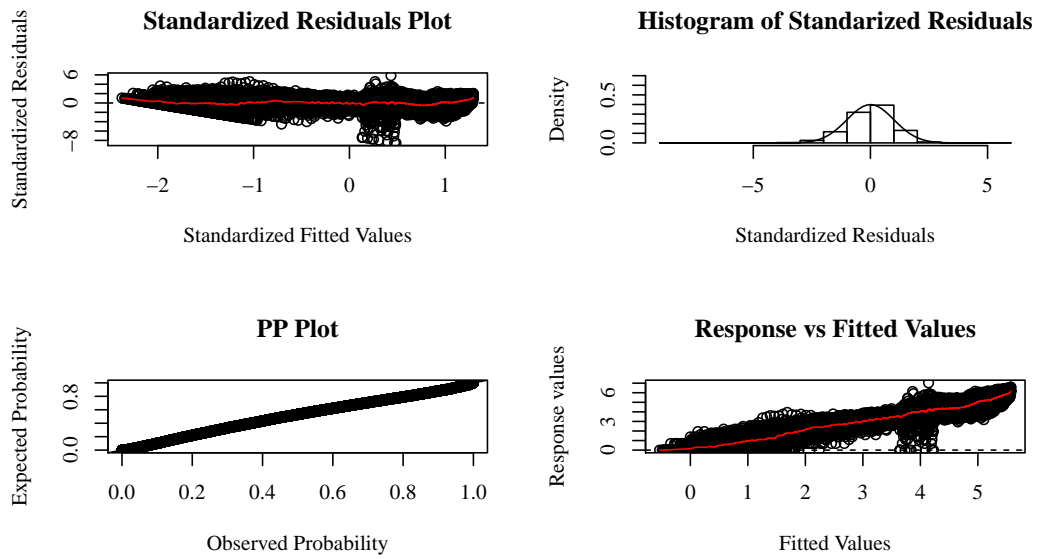
PP Plot

Response vs Fitted Values

Figure 13: Diagnostic Plot Linear Model without Interaction (LMNI) by Age.

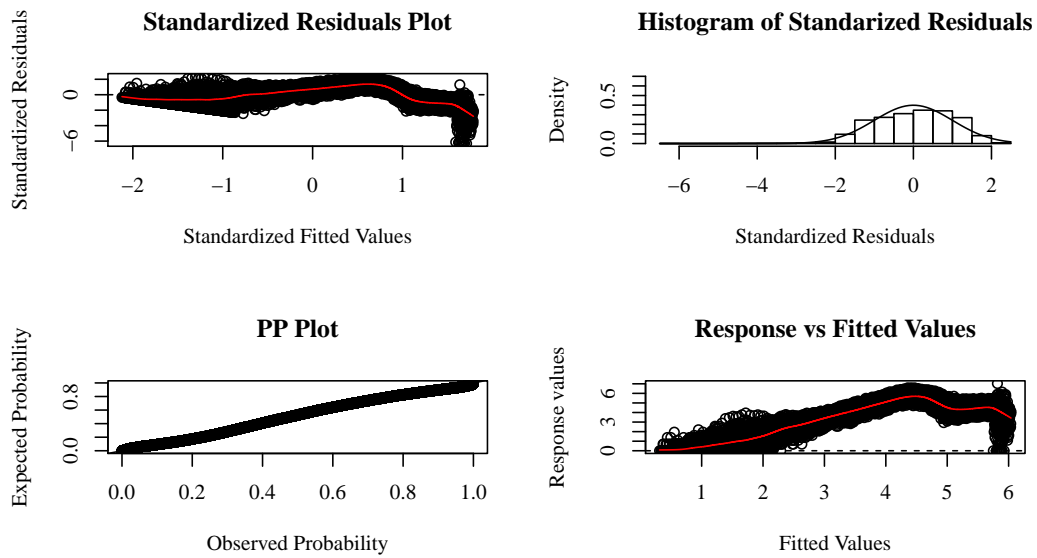Figure 14: Diagnostic Plot Additive Model without Interaction (AMNI) by Age.

Figure 15: Diagnostic Plot Linear Model with Interaction (LMI) by Age.

References

De Boor, C. (1978). *A practical guide to splines* (Vol. 27). Springer-Verlag New York.

Duchon, J. (1977). Splines minimizing rotation-invariant semi-norms in sobolev spaces. In *Constructive theory of functions of several variables* (pp. 85–100). Springer.

Geweke, J., & Whiteman, C. (2006). Bayesian forecasting. *Handbook of economic forecasting*, *1*, 3–80.

Wood, S. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(1), 95–114.

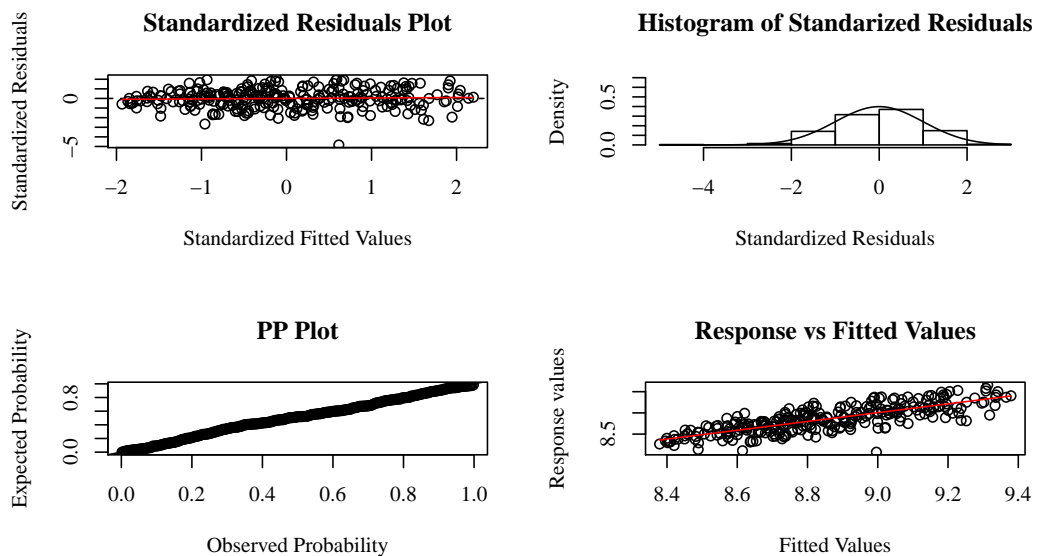Wood, S. (2006). *Generalized additive models: an introduction with r.* CRC press.

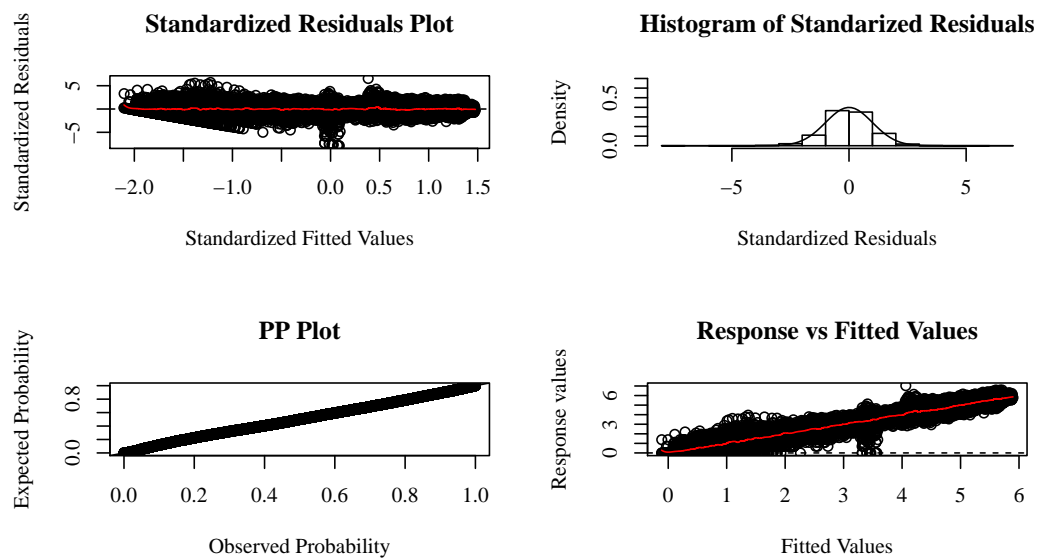Figure 16: Diagnostic Plot Additive Model with Interaction (AMI) by Age.

Figure 17: Diagnostic Plot Nonparametric Model with Interaction (NPM) by Age.