Y데이터의 경우 120달(2009년 1월 ~ 2018년 12월)동안의 한국에 입국한 국적별 외국인 수를 나타내었고, X데이터의 경우 120달 동안의 각 나라의 한국 원화 기준 환율을 나타낸 것이다. 나라는 한글 순으로 [가나, 가봉, … , 헝가리, 호주, 홍콩] 174개국으로 구성했으며, 시간의 흐름에 따라 데이터를 나열했다. R 기반으로 데이터를 정제했으며, 분석 또한 R로 이루어 졌다.

X, Y 데이터 모두 리스트화를 거쳤다. Y데이터가 허들모델이라는 가정으로 각 행마다 0이 얼마나 포함되어 있는지 알아보았다. 그 결과 32, 71, 94, 100, 104, 112, 113행에는 0을 포함하지 않아서 허들모델이나 zero inflated 방법을 이용하여 모델을 적합할 수 없었다. 그래서 우리는 Goodness of Fit(GoF)를 이용하여 일반화 선형모형의 적합도를 검정해보았다.

그 결과 포아송 GoF는 모두 0으로 나왔으며, 음이항분포 GoF는 낮은 값을 보였다. 즉, 포아송분포를 사용하였을 때 과대산포가 발생하므로, 음이항분포를 이용하여 모형적합을 시도했다.

X의 범위를 이용하여 커널 메트릭스를 생성하였으며, GCV 벡터를 찾기 위해 범위에서 적절한 를 찾아냈다. 이 때 구한 를 이용하여 의 벡터를 리스트화했고, 이 때 논문의 방법인 EM알고리즘을 통해 multilevel spline 방법으로 구한 의 벡터(120x1)를 구할 수 있다. Naïve한 방법 역시 같은 방식으로 구한다.

