Hierarchical Time Series Models for the Onset of COVID-19 in Northeast Indiana

Peter P. Dragnev*

*Department of Mathematical Sciences, Purdue University Fort Wayne
Supervisor: Alessandro Maria Selvitella*

PURDUE UNIVERSITY.
FORT WAYNE

Northeast Indiana's Pandemic Challenges

In 2020, the world faced a once-in-a-century pandemic. As the COVID-19 novel virus spread around the globe in the early Winter and Spring months of 2020, communities in Northeast Indiana were also affected – causing economic and health challenges that still remain today. Many organizations transitioned to remote work. Schools offered both in-person, virtual and blended options – challenging teachers and students to adapt their instruction and learning methods. Some companies laid off large portions of their workforce or completely shut down and and sold off, due to lost business. Others highly profited from their e-commerce business models. Many bars and restaurants struggled to stay in business – adapting to take-out and delivery service. Hospitals struggled to keep up with the influx of COVID patients, while still trying to provide routine, non-COVID health services. In addition to affecting the economy, COVID has also had an effect on the physical, emotional and mental health of many individuals in our community. Although life in the region is slowly returning to normal, many challenges still remain.

The Problem

Modeling and predicting the spread of the virus is essential to state and local government authorities and public health officials in making decisions about closures and restrictions to control case numbers. This project models the time evolution of the original COVID-19 outbreak in our region using historical data, which may be useful as future variants and strains arise and affect various communities around the world, including those in Northeast Indiana – as demonstrated by the recent Delta surge.

The initial infectious period, the spread can be exhibited by the population growth model $\frac{dy(t)}{dt} = \beta\, y(t)$. The solution of this growth model is $y(t) = k\, e^{\beta\, t}$, where $k = y(0)$ is the initial value at time $t = 0$. This fact leads to a time series model with exponential deterministic trend, up to an error term. Note that this model is a first order approximation at the initial time of many of the equations for the infected compartment in the systems of ordinary differential equations defining some popular compartmental models, such as the SIR (Susceptible-Infectious-Removed) model [1]. In such a model, one has that around time $t = 0$:

$$\frac{dy}{dt} = \left(R_0 \frac{S}{N} - 1\right) \gamma y,$$

and so $\beta = \left(R_0 \frac{S}{N} - 1\right)\gamma$ carries important information about epidemiological parameters such as the typical time until removal γ and the basic reproduction number R_0 (Here S and N are the initial susceptible and total population, respectively). In general though, the fitting of this model suffers from residuals which do contain some further correlation structure, such as a stochastic trend. For this reason, in this project, we allow this error term to be autoregressive of order 1 and verify how giving this extra structure to the error affects the fit to data relevant to the beginning of the COVID-19 pandemic in the counties of Northeast Indiana.

Methods, Data and Scope

This project applies linear regression analysis between log of the number of cases y and time t: $\log\, \widehat{y}_t = \beta_0 + \beta_1 t + \varepsilon_t$. It also analyzes the error ε_t via an autoregressive process of order one AR(1) for linear relationship: $\varepsilon_t = c + \alpha\varepsilon_{t-1} + \eta_t$, where $\eta_t \sim N(0, \sigma^2)$ is a white noise process [3]. The data used in this project comes from the Indiana State Department of Health (ISDH), available on the Indiana Data Hub [2]. The analysis considers the Northeast Indiana region and each of its 11 counties, during the initial 3-week periods of widespread infection in the respective counties.

Linear Regression Analysis

The table below summarizes the linear regression results for the Northeast Indiana region, as well as each individual county:

COVID Onset in Northeast Indiana ($\log\, \widehat{y}_t = \beta_0 + \beta_1 t + \varepsilon_t$)				
Region / County	Onset Date Range	Lin. Reg. Coefficients	R^2	β_1^* Conf. Interval
NE Indiana	3.22.20 - 4.11.20	β_0^*=1.3507, β_1^*=0.2245	0.8862	(0.1858,0.2631)
Adams	8.04.20 - 8.24.20	β_0^*=2.0241, β_1^*=0.1369	0.7929	(0.1033,0.1705)
Allen	3.25.20 - 4.14.20	β_0^*=2.0128, β_1^*=0.1691	0.7861	(0.1267,0.2114)
De Kalb	8.04.20 - 8.24.20	β_0^*=1.1726, β_1^*=0.1661	0.8390	(0.1311,0.2010)
Huntington	8.25.20 - 9.14.20	β_0^*=1.2876, β_1^*=0.1435	0.8113	(0.1103,0.1768)
Kosciusko	5.20.20 - 6.09.20	β_0^*=1.8139, β_1^*=0.1948	0.8210	(0.1511,0.2385)
Lagrange	5.25.20 - 6.14.20	β_0^*=1.3071, β_1^*=0.2265	0.8835	(0.1870,0.2660)
Noble	4.12.20 - 5.02.20	β_0^*=1.4875, β_1^*=0.1517	0.7450	(0.1091,0.1943)
Steuben	7.14.20 - 8.03.20	β_0^*=1.4717, β_1^*=0.1529	0.7602	(0.1116,0.1941)
Wabash	9.24.20 - 10.14.20	β_0^*=1.3650, β_1^*=0.1555	0.8586	(0.1252,0.1858)
Wells	9.24.20 - 10.14.20	β_0^*=2.2965, β_1^*=0.1190	0.9915	(0.1137,0.1243)
Whitley	9.18.20 - 10.08.20	β_0^*=1.4481, β_1^*=0.1769	0.8912	(0.1473,0.2066)

The 3-week date ranges – reflecting the initial periods of widespread infection in respective Northeast Indiana counties – were selected to avoid stretches of zero daily cases and produce more accurate infection models. Confidence intervals for regression coefficients β_1^* with significance level $\alpha = 0.05$ are also produced.

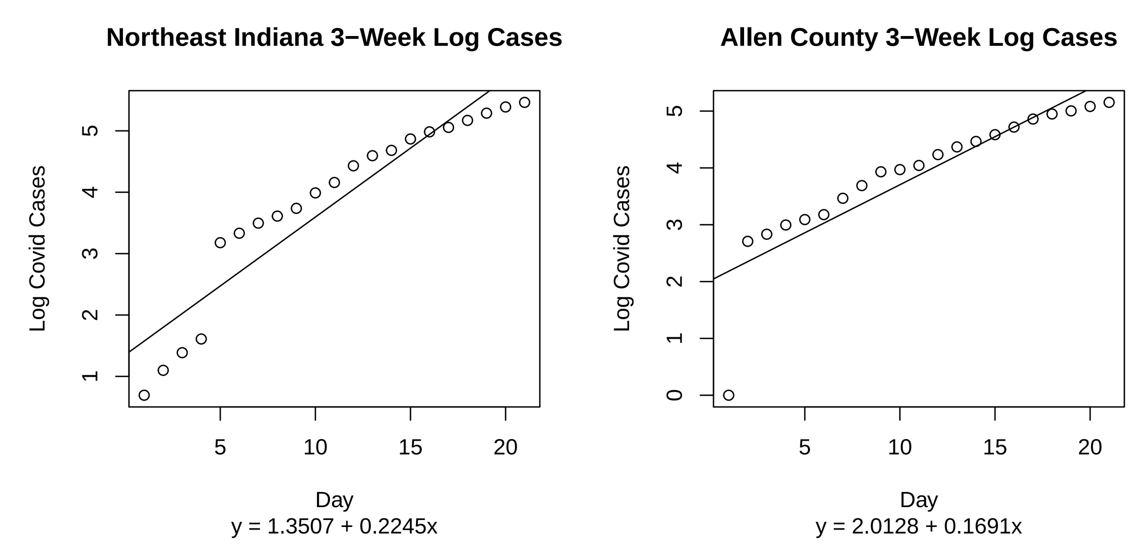

Fig. 1: Northeast Indiana Region and Allen County 3-Week Log Cases

During the onset of COVID-19 in spring of 2020, Allen County, the most populous county in the region, was the earliest to be impacted by widespread infection. Many of the other counties in the region did not see their first major spread until the late summer season. Figure 1 shows log cases for Northeast Indiana and Allen County. The jump in regional log cases on day 5 of the Northeast Indiana model, is due an influx of positive cases in Allen County that occurred on March 26, 2020 - displayed on day 2 of of the Allen County model.

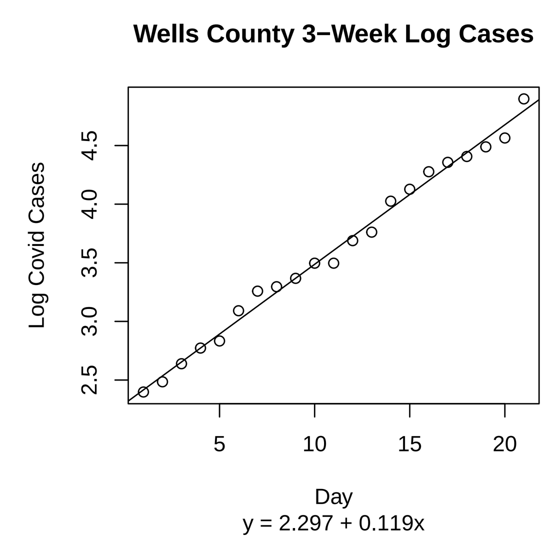

Fig. 2: Wells County 3-Week Log Cases

The coefficients of determination R^2 for all localities, as well as the region as a whole, are 0.75 or higher, which suggests that the respective models are a good fit for the initial spread in each individual county. For Wells county in particular, the model is an almost perfect fit, with an R^2 value of 0.9915 - see Figure 2.

Error Analysis

The table below summarizes results of the error ε_t via an autoregressive process of order one AR(1) for the Northeast Indiana region and each individual county:

COVID Onset in Northeast Indiana ($\varepsilon_t = c + \alpha\varepsilon_{t-1} + \eta_t$, where $\eta_t \sim N(0, \sigma^2)$)			
Region / County	Onset Date Range	Error Coefficients	σ^2
NE Indiana	3.22.20 - 4.11.20	c=-0.2322, α=0.8272	0.0913
Adams	8.04.20 - 8.24.20	c=-0.4618, α=0.9035	0.0784
Allen	3.25.20 - 4.14.20	c=-0.0368, α=0.2347	0.2817
De Kalb	8.04.20 - 8.24.20	c=-0.4701, α=0.9303	0.0543
Huntington	8.25.20 - 9.14.20	c=-0.3477, α=0.8644	0.0912
Kosciusko	5.20.20 - 6.09.20	c=-0.5358, α=0.8773	0.1637
Lagrange	5.25.20 - 6.14.20	c=-0.6311, α=0.9349	0.0695
Noble	4.12.20 - 5.02.20	c=-0.4172, α=0.8770	0.1226
Steuben	7.14.20 - 8.03.20	c=-0.5779, α=0.9176	0.0925
Wabash	9.24.20 - 10.14.20	c=-0.2227, α=0.7826	0.1086
Wells	9.24.20 - 10.14.20	c= 0.0012, α=0.2338	0.0042
Whitley	9.18.20 - 10.08.20	c=-0.3158, α=0.8967	0.0448

The σ^2 value of 0.0042 for Wells county is notably lower than the other counties and the region, suggesting a very narrow normal curve of error distribution. This reflects the high value R^2 of 0.9915 in the regression analysis table.

Model Testing

The table below compares the predicted log case values for day 22 to the actual recorded log case values:

3-Week Predictive Model Testing ($\log\, \widehat{y}_t = \beta_0 + \beta_1 t + \varepsilon_t$)			
Region / County	Actual Log Cases	Predicted Log Cases	Error
NE Indiana	$\log\, y_{22}$ = 5.5175	$\log\, \widehat{y}_{22}$ = 6.2897	ε_{22} = -0.7722
Adams	$\log\, y_{22}$ = 4.4427	$\log\, \widehat{y}_{22}$ = 5.0359	ε_{22} = -0.5932
Allen	$\log\, y_{22}$ = 5.3845	$\log\, \widehat{y}_{22}$ = 5.7330	ε_{22} = -0.3485
De Kalb	$\log\, y_{22}$ = 4.4427	$\log\, \widehat{y}_{22}$ = 4.8268	ε_{22} = -0.3841
Huntington	$\log\, y_{22}$ = 3.9512	$\log\, \widehat{y}_{22}$ = 4.4446	ε_{22} = -0.4934
Kosciusko	$\log\, y_{22}$ = 5.3660	$\log\, \widehat{y}_{22}$ = 6.0995	ε_{22} = -0.7335
Lagrange	$\log\, y_{22}$ = 5.4765	$\log\, \widehat{y}_{22}$ = 6.2901	ε_{22} = -0.8136
Noble	$\log\, y_{22}$ = 4.2627	$\log\, \widehat{y}_{22}$ = 4.8249	ε_{22} = -0.5622
Steuben	$\log\, y_{22}$ = 4.0943	$\log\, \widehat{y}_{22}$ = 4.8355	ε_{22} = -0.7412
Wabash	$\log\, y_{22}$ = 4.4773	$\log\, \widehat{y}_{22}$ = 4.7860	ε_{22} = -0.3087
Wells	$\log\, y_{22}$ = 4.9698	$\log\, \widehat{y}_{22}$ = 4.9145	ε_{22} = 0.0553
Whitley	$\log\, y_{22}$ = 4.7621	$\log\, \widehat{y}_{22}$ = 5.3399	ε_{22} = -0.5777

With the exception of Wells county, all other residuals for day 22 are negative – indicating a possible slowdown in the exponential growth of cases after 3 weeks.

References

[1] Fred Brauer, Carlos Castillo-Chavez, and Zhilan Feng. *Mathematical Models in Epidemiology*. Vol. 69. Springer, 2019.

[2] ISDH Indiana Department of Health. *COVID 19 County-wide Test, Case, and Death Trends.* 2021. URL: `https://hub.mph.in.gov/dataset/covid-19-county-wide-test-case-and-death-trends` (visited on 10/10/2021).

[3] Wikipedia. *Autoregressive Model, Example: An AR(1) process.* 2021. URL: `https://en.wikipedia.org/wiki/Autoregressive_model#Example:_An_AR(1)_process` (visited on 10/25/2021).