Interrupted Time Series with Autoregressive Moving Average Models and Hidden Markov Models

Chaoyi Zheng

Advisor: James Li, Ph.D.

Department of Biostatistics, Bioinformatics, and Biomathematics,

Georgetown University, Washington DC

January 2017

Abstract

Motivation: Time series data often violate the assumptions of linear regression. Interrupted time series (ITS) analysis requires more complex models to generate reliable inference. Though popular in machine learning, the applicability of hidden Markov models in ITS is largely unexplored.

Objectives and Methods: This study aims at finding appropriate models to describe the monthly total number of cancer surgery cases in the New York State between 1997-2006. It attempts to perform ITS analysis using autoregressive moving average models (ARMA) and hidden Markov models (HMM), assuming a segmented regression framework. The two approaches were compared on predictive precision and ability to estimate intervention effects.

Results: The series exhibited clear seasonal pattern and autocorrelation not accounted for by simple segmented regression. Both ARMA and HMM were able to arrive at models adequate to describe the data; they also both greatly improved the predictive accuracy compared with the simple segmented regression. An extended HMM model with additional structure for direct autocorrelation among observations had greater predictive accuracy than the ARMA model (SSE for predicting the last year of data was 108,270 vs 125,391). However, current tools in R were not able to generate reliable estimates of the intervention effects from HMM. The ARMA model concluded that the intervention did not significantly affect the outcome.

Conclusions: Both ARMA and HMM are significant improvements over segmented linear regression. While HMM performed much better in the prediction task, ARMA has been more thoroughly developed to estimate intervention effects. HMM's utility in ITS studies is worth further exploration.

1. Background

In health policy research, it is often of interest to evaluate the effect of a policy intervention. One widely used design for this type of study is interrupted time series (ITS), where the outcome of interest is measured at equally-spaced time points (time series) and analyzed with respect to changes after the intervention.(Kontopantelis et al. 2015) Unlike cross-sectional data, time series data have distinct characteristics that frequently violate the assumptions of linear regression.

As summarized elsewhere, major characteristics in time series data include trending, periodicity, structural breaks, and autocorrelation.(Pickup 2014) These characteristics violate assumptions of zero conditional mean (exogeneity) and independent errors. The former leads to biased estimates of regression coefficients, and the latter can distort estimates of their standard errors (Kutner et al. 2004). Both undermine the validity of inferences drawn from the regression.

One popular type of models in ITS studies is segmented regression that incorporates an indicator for time before or after the intervention as a covariate (Wagner et al. 2002; Bernal, Cummins, and Gasparrini 2016). Typically, these models also include a term for pre-existing time trend for the series as well as an interaction term between the indicator and time. As a result, these three terms represent the change in intercept after intervention, the baseline slope, and the change in slope after the intervention, respectively. The segmented regression models are useful because their baseline slope term can account for trending in the series if the trend is linear, while the other two terms model the structural break of interest, i.e. the policy intervention.

However, the simple segmented regression models do not account for periodicity and autocorrelation in the series, and are therefore often inadequate. One approach to handle this issue is to extend the traditional regression models to include terms corresponding to these additional processes, such as the autoregressive integrated moving average (ARIMA) models.

Another approach to this problem is to account for these patterns through structures unlike regression. The unique dependency structure in a Markov chain makes it suitable for the task. Hidden Markov models (HMM) takes advantage of this feature (Zucchini and MacDonald 2009). Despite its vast popularity in many areas of machine learning, HMM has not been widely used in policy research. Particularly, the author has found no published examples of its application in intervention studies. Methods to incorporate the segmented regression model with the HMM framework have not been thoroughly developed. This study makes an attempt to this direction.

This study aims at addressing these inadequacies by enriching the skeleton of segmented regression through these two distinct approaches. The series used for this study are the monthly total number of inpatient cancer surgery cases performed in the New York State between 1997 and 2006, and the policy intervention of interest is New York state's expansion of Medicaid in October, 2001. The aims of this study are twofold – 1) to evaluate the performance of ARMA and HMM on these data in terms of adequacy and predictive accuracy, and 2) to compare the two approaches' estimates of the effects of a policy intervention.

2. Methods

2.1 Inferences from Time Series Data

Let Y_t denote a time series

$$Y_t = \{y_1, y_2, \ldots, y_T\}; \quad T = number\ of\ time\ points \tag{1}$$

Unlike cross-sectional data, time series data are not produced by repeated random sampling from a population; rather, a time series is a single realization of a data-generating process, with a single draw at each time point (or time period). While analysis on cross-sectional data usually makes inferences about the population from which the sample is drawn, in time series, the inferences are made regarding the data-generating process (Pickup 2014).

In addition to the assumptions made by ordinary least squares (OLS) regression, time series regression also requires the series to be covariance stationary to achieve unbiasedness for its estimates of parameters and their variances (Pickup 2014). A stochastic process is considered covariance stationary if $E(y_t)$ is constant, $Var(y_t)$ is constant, and $Cov(y_t, y_{t+h})$ is constant given the same h. Covariance stationarity is a weaker form of stationarity, which requires $\{y_k, y_{k+1}, \ldots, y_{k+h}\}$ to be identically distributed for all k when given h (Pickup 2014).

2.2 Autocorrelation

Relative positions among individual data points in a time series is described as "lags" and "leads". For instance, y_{t-1} is the first lag of y_t, while y_{t+1} is the first lead of y_t. Autocorrelation describes the linear correlation between a realization with its lags or leads (Pickup 2014). For the series Y_t, its sth autocorrelation coefficient is

$$\rho_s \equiv \frac{E((y_t - \mu_y)(y_{t-s} - \mu_y))}{E(y_t - \mu_y)^2} \tag{2}$$

Under stationarity assumptions, ρ_s can be estimated as

$$\widehat{\rho_s} \equiv \frac{\sum_{t=s+1}^{T}((y_t - \bar{y})(y_{t-s} - \bar{y}))}{\sum_{t=1}^{T}(y_t - \bar{y})^2} \tag{3}$$

The autocorrelation function (ACF) is the function projecting s to ρ_s. Its analogue, the partial autocorrelation function (PACF), substitutes correlation with partial correlation that controls for all lags between y_t and y_{t-s}. Trending, periodicity and dependent observations can all be reflected in ACF and PACF. Therefore, these functions are crucial diagnostic tools in the ARIMA model building process.

2.3 ARIMA Models

A typical segmented regression model can be written as

$$y_t = \beta_0 + \beta_1 \times time + \beta_2 \times I_t + \beta_3 \times time \times I_t + \epsilon_t; \quad \epsilon_t \sim i.i.d. \, N(0, \sigma^2) \qquad (4)$$

$$I_t: indicator \; for \; policy \; intervention, I_t = \begin{cases} 0, & before \; the \; intervention \\ 1, & after \; the \; intervention \end{cases}$$

$$time: counter \; of \; time \; period, takes \; values \; of \; 1,2,3, \dots$$

As discussed earlier, β_1 in this model represents the baseline linear trend in the series prior to the intervention; β_2 represents the intercept change after the intervention and β_3 the slope change after the intervention (Wagner et al. 2002; Kontopantelis et al. 2015). The latter two terms together model the potential structural break caused by the intervention, assuming homoscedasticity throughout the study period. The segmented regression model has the advantage of being easy to interpret, and can be considered a quasi-experimental design that adjust for the baseline trend in the series when evaluating the effects of an intervention.

The autoregressive integrated moving average (ARIMA) models can enrich the simple segmented regression model by incorporating components suitable to model the underlying structures of a time series: autoregressive (AR) component, integrative component (I), and moving average component (MA). A model with all these three components is an ARIMA model, while a model with both AR and MA components but no I component is an ARMA model, etc.

The integrative component is typically used to model instable series that do not converge to an equilibrium. For the purpose of this study, this component is omitted as the instability in health policy series can often be taken care of by linear adjustment of time trend (i.e. β_1 in (4)). Therefore, the ARIMA models discussed in this study are ARMA models.

Figure 1 represents the data-generating process modeled by ARMA. It starts with a random shock ϵ_t (a.k.a. error term, stochastic component) from a homoscedastic white noise process with no structure nor serial correlation. The random shock then undergoes AR processes where an observation draws on its past value and/or MA processes where the stochastic component draws on its past value (McDowall, McCleary, and Meidinger 1980). To build an ARMA model is to reverse the data-generating process by parsing out the AR and/or MA processes to convert the observed series y_t back to ϵ_t, a stationary series, so that inferences can be made from the series in a fashion

similar to linear regressions.

Figure 1. The Autoregressive Moving Average (ARMA) Process.

An ARMA model with AR process to the order of p and MA process to the order of q, denoted as ARMA(p, q), can be written as below (2014):

$$\Phi(B)\tilde{y}_t = \Theta(B)\epsilon_t \tag{5}$$

$\tilde{y}_t$	Deviation from equilibrium; $y_t - \widetilde{\mu_y}$
$\widetilde{\mu_y}$	Covariate-adjusted μ_y; In this study, $\widetilde{\mu_y} = \mu_y + \beta_1 \times time + \beta_2 \times I_t + \beta_3 \times time \times I_t$
B	The backshift operator; $B(y_t) = y_{t-1}$; $B^i(y_t) = y_{t-i}$
$\Phi(B)$	The autoregressive operator; $1 - \sum_{i=1}^{p} \phi_i B^i$
$\Theta(B)$	The moving average operator; $1 - \sum_{j=1}^{q} \theta_j B^j$

In expanded form:

$$\tilde{y}_t = \sum_{i=1}^{p} \phi_i \tilde{y}_{t-i} + \epsilon_t - \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \tag{6}$$

2.3.1 The Box-Jenkins Approach for Model Selection in ARIMA

The Box-Jenkins approach is the most widely-used approach to specify an appropriate ARIMA model. It starts with a stationary data series, and iteratively alternates between identifying the data-generating mechanism and estimating the data model. Formally, this approach involves three stages: 1) identification, 2) estimation, and 3) diagnostic checking (Pickup 2014).

2.3.1.1 Identifying a Plausible Model

The Box-Jenkins approach utilizes ACF and PACF plots to identify possible AR and/or MA processes by matching the plots with theoretical patterns generated by corresponding processes. An example of the guidelines for this procedure is exhibited in Table 1. It is worth noting that, according to the Wold's theorem, a finite AR process can be written as a longer or infinite MA process. For model parsimony, finite processes are preferred to represent the data.

Table 1. Guidelines for Identifying Autoregressive and Moving Average Processes (after Pickup, 2015[*]).

Process	ACF	PACF
White noise	ACF = 0 at all lags ≥ 1.	PACF = 0 at all lags ≥ 1.
AR(p)	Decays toward zero.	Spikes at p[th] lag. All PACF = 0 at lags greater than p.
MA(q)	Spike at s = q and ACF = 0 for all lags other than q.	Decay toward zero.
ARMA (p, q)	Decay beginning at lag q.	Decay beginning after lag p.

*Pickup, Mark. Introduction to Time Series Analysis (Quantitative Applications in the Social Sciences) (p. 130). SAGE Publications. Kindle Edition.

2.3.1.2 Estimating the Model

In this step, the model identified in the previous step is fitted. When no missing value is present, ARIMA models can be fitted with both maximum likelihood and least squares. In this study, these models are fitted by arima functions in base R using conditional sum-of-squares to find initial values then maximum likelihood to optimize ("R: ARIMA Modelling of Time Series" 2017).

2.3.1.3 Check the Model Fit Using Diagnostics

To verify whether the fitted model sufficiently converted the observed series back into random shocks, a white noise test such as the Portmanteau Q test was performed (Pickup 2014). This test is also referred to as the Ljung-Box test. The ACF and PACF plots can also be used to check whether the residuals are a white noise (see Table 1). If the model does not pass the white noise test, it should be revised and the Box-Jenkins procedure iterated.

$$Q = T(T-2) \sum_{s=1}^{p} \frac{\widehat{\rho_s}^2}{T-s} \tag{7}$$

Under $H_0: \rho_1, \rho_2, \ldots, \rho_p = 0, Q \sim \chi_p^2$.

Other diagnostic tests can be performed to check the exogeneity assumption and the homoscedasticity assumption. To compare and select among adequate models, criteria such as AIC, BIC can be used. Likelihood-ratio tests can be used to compare nesting models (Pickup 2014).

2.4 Hidden Markov Models

The second approach used in this study to model time series data is hidden Markov models (HMM). They are fundamentally different from regression-based methods like segmented regression and ARIMA. We start by offering a brief description of these models.

HMMs are a class of mixture models that describe the observed distribution as a mixture of several underlying source distributions. Specifically, HMM assumes the mixture to be dependent on a latent Markov chain that switches among a finite number of states. Each source distribution (referred to as component) uniquely corresponds to one state of the Markov chain. The probability for a given node c_t in the chain to assume state i is the weight of the corresponding component in the mixture distribution at time t. Each node (c_t) in the Markov chain corresponds to a unique time point t. To put mathematically, for discrete-valued observations Y_t, assuming the Markov chain has m states,

$$P(y_t = y) = \sum_{i=1}^{m} P(c_t = i) P(y_t = y | c_t = i) \tag{8}$$

Description of continuously-distributed y_t is analogous (Zucchini and MacDonald 2009).

In a first-order discrete-time Markov chain C_t, the state of the chain at a given node is dependent on the state of the previous node; the conditional distribution of node c_t given the state of c_{t-1} is independent of all previous nodes (Zucchini and MacDonald 2009). The transition matrix $\Gamma \{\gamma_{ij}\}$ describes the probability of the transitions. The chain is referred to as homogeneous if Γ does not vary with t.

$$P(c_t|C^{t-1}) = P(c_t|c_{t-1}) \qquad (9)$$

$$P(c_{t+1} = j) = \gamma_{ij}P(c_t = i) \qquad (10)$$

C^{t-1} Vector for states assumed by the Markov chain from time point 1 to t-1

γ_{ij} Elements of the $m \times m$ transition matrix Γ.

As shown in Figure 2(a), HMMs are uniquely suitable for modeling time series data with autocorrelated observations in that the latent Markov chain structure has the inherent feature of serial dependence. The HMM structure also has extensive flexibility in terms of introducing covariates and adding other dependency patterns (Zucchini and MacDonald 2009). In this study, extended models incorporating additional dependence among the observations (Figure 2(b)) were fitted and compared with the basic model in Figure 2(a).
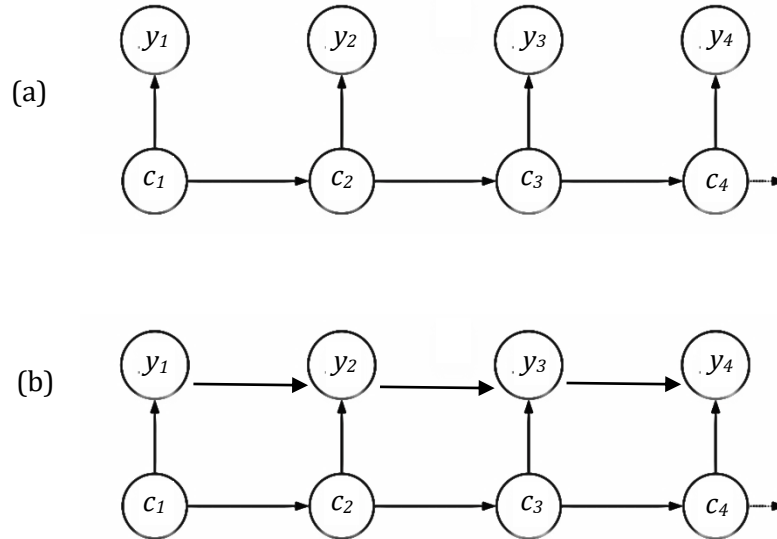

Figure 2. Directed Graph Representations of Hidden Markov Models. (a) Basic model; (b) extended model with autocorrelated observation process.

2.4.1 Specification of Component Distributions in HMM

In this study, although the observations are count data, Gaussian distribution was chosen over Poisson distribution to model the state-dependent observations to ensure the interpretability of the model, which can be complicated by the necessity to scale and offset the data and potential over-dispersion issues if Poisson models were used. The resulting models are referred to as HMM-Gaussian models. The distributions of observations given $c_t = i$ are modeled through linear models.

$$y_{i,t} = \beta_{0,i} + \sum_k \beta_{k,i} x_{k,i,t} + \beta_{s,i} \sin\frac{2\pi t}{12} + \beta_{c,i} \cos\frac{2\pi t}{12} + \epsilon_t \tag{11}$$

$\sum_k \beta_k x_{k,t}$	Sum of covariate terms (k = 1, 2, 3) for segmented regression as specified as in (4).
$\sin\frac{2\pi t}{12}, \cos\frac{2\pi t}{12}$	First-order Fourier terms for seasonality adjustment (Rob J Hyndman 2014). Fourier terms were chosen over 12th-order autoregression to avoid losing 12 observations. This issue is overcome in ARIMA by advanced estimation methods, but is not readily resolved in these specific HMM.

In the extended HMM models with autocorrelated observation process, the state-dependent models were enriched to incorporate autoregressive terms:

$$y_{i,t} = \beta_{0,i} + \sum_k \beta_{k,i} x_{k,i,t} + \beta_{s,i} \sin\frac{2\pi t}{12} + \beta_{c,i} \cos\frac{2\pi t}{12} + \sum_{i=1}^{p} \beta_i x_{t-i} + \epsilon_t \tag{12}$$

2.4.2 Estimation of HMM

One of the mathematically attractive features of HMM is that its likelihood is straightforward to compute. The likelihood of an HMM can be written as (Zucchini and MacDonald 2009):

$$L_T = \Delta P(y_1)\Gamma P(y_2)\Gamma P(y_3)\ldots\Gamma P(y_T)1' \tag{13}$$

$P(y_t)$	Probability of $Y_t = y_t$ at time t; $diag\{p_1(y_t), p_2(y_t), \ldots, p_m(y_t)\}$
Δ	Starting distribution of the Markov chain; distribution of C_t
Γ	Transition matrix of the Markov chain

The expectation-maximization (EM) algorithm is most often used to find maximum-likelihood estimates of HMM. The Baum-Welch EM algorithm assumes the Markov chain to be homogeneous but not necessarily stationary (i.e. the distribution of states reaches an equilibrium unaltered by transitions). The complete-data log-likelihood can be decomposed as the sum of three terms, each regarding one set of parameters in HMM: the initial distribution Δ, the transition matrix Γ, and the parameters of the state-dependent component distributions. In Baum-Welch EM, the expectation of distribution of the Markov chain are maximized regarding these parameters (Zucchini and MacDonald 2009). The number of states possible for the chain (m) needs to be pre-specified.

$$\log P(y^{(T)}, c^{(T)}) = \sum_{j=1}^{m} u_j(1)\log\delta_j + \sum_{j=1}^{m}\sum_{i=1}^{m}\left(\sum_{t=2}^{T} v_{jk}(t)\right)\log\gamma_{jk} + \sum_{j=1}^{m}\sum_{t=1}^{T} u_j(t)\log p_j(x_t) \tag{14}$$

$$= L_1\left(u_j(1)\right) + L_2(\Gamma) + L_3\left(p_j(x_t)\right)$$

$$u_j(t) \qquad\qquad \text{Indicator that } c_t = j$$
$$v_{jk}(t) \qquad\qquad \text{Indicator that } c_{t-1} = j \text{ and } c_t = k$$

However, for the purpose of quantifying the effects of an intervention through the $\beta_{k,i}$ terms in (11) and (12), it would be useful to add constraints to the model fitting to make the coefficients describing the post-intervention changes equal across all latent states for the same k (i.e. let $\beta_{k,1} \equiv \beta_{k,2} \equiv \cdots \equiv \beta_{k,m} \equiv \beta_k$, for k=2 or 3). That is, while the observations are state-dependent, the effect of the intervention are assumed to be state-invariant.

$$y_{i,t} = \beta_{0,i} + \beta_{1,i} \times time + \beta_2 \times I_t + \beta_3 \times time \times I_t + \beta_{s,i} \sin\frac{2\pi t}{12} + \beta_{c,i} \cos\frac{2\pi t}{12} + \sum_{i=1}^{p} \beta_i x_{t-i} + \epsilon_t \qquad (15)$$

In this case, EM algorithm can produce bad estimates for the parameters. The R package **depmixS4** uses non-linear optimization with Lagrange multiplier for this scenario. Because good starting values is required for this method, the models were first fitted without parameter constraints (as in (12)); the parameter estimates from (12) were then used as a starting point to fit (15). This area of the R package is still being improved; as of now, estimates of the variance of the parameter estimators are lacking (Visser and Speekenbrink 2016).

2.5 Model Comparison

Within both the ARMA and HMM approach, multiple models were fitted. ARMA models were compared based on their adequacy as a part of the Box-Jenkins model building process. HMM models were compared on AIC and the one(s) with smallest AIC were chosen. Their adequacy was examined through pseudo-residuals. If an HMM is adequate, its pseudo-residuals should follow a uniform distribution $U(0,1)$. To detect outliers more easily, these pseudo-residuals are transformed into normal pseudo-residuals and visualized via Q-Q plots. It should be noted that these pseudo-residuals should not be assumed to be independent (Zucchini and MacDonald 2009). In this study, the normal pseudo-residuals were calculated with the **HiddenMarkov** R package (Harte 2016).

The best-performing and adequate model(s) from each approach were then compared on their ability to predict future series. This evaluation was done through training the respective models with data from the first 9 years and testing the them on their predictive accuracy using the last year of data. The testing observations were held back from all model fitting. Accuracy was measured as the sum of square errors (SSE) of the predictions.

Finally, estimates of β_2 and β_3 were compared among simple segmented regression, ARMA and HMM. The constrained HMM model (15) was fitted only for this purpose, while fit statistics from (12) were used for all model comparisons.

3. Experiment

The time series used in this study concerns the monthly total number of inpatient cancer surgeries performed in the New York State. It's a 10-year series with 12 data points per year; $T = 120$. This data comes from the State Inpatient Database of New York, which were available through the Health Care Utilization Project under Agency for Health Quality Research. The policy intervention of interest is the expansion of New York State's Medicaid program in October, 2001. In theory, a policy that increases access to healthcare could cause increase in utilization of healthcare. Therefore, is of interest whether this intervention caused an increase in the series.

Over the study period, the series had an overall mean of 1,324 (cases/month) and standard error of 104, IQR was 1,255-1,398. Examination of the distribution of observations over time clearly suggests an upward trend (Figure 3Figure 3Figure 3Figure 3Figure 3Figure 3).

3.1 ARMA

Figure 4 shows the improvement in model adequacy from null model to fully-fitted seasonal ARMA models. The dotted lines in each panel represents the significance thresholds for ACF or PACF values. In the null model, the slowly decaying positive ACF indicates instable series (Figure 4a). This pattern was eliminated after adjusting for time trend in the segmented regression model (Figure 4b). However, it is evident from the ACF plot of segmented regression residuals that the observations are serially correlated at the first lag. Similarly, a 12th ACF as high as 0.4 likely suggests a strong seasonal pattern, where the number of cancer surgery in a specific month (e.g. January) is correlated across different years. The PACF has more complex patterns and is harder to interpret.

Comparing Figure 4c and 4d with 4b, it can be seen that the two ARMA models have successfully eliminated autocorrelation in the model residuals. The model in 4c is $AR(1) \times (1)_{12}$. Put in the form of (5), $\Phi(B) = (1 - \phi_1 B)(1 - \phi_{1 \times 12} B^{12})$ and $\Theta(B) = 1$. The model in 4d is $ARMA(0,1) \times (1,0)_{12}$, or $\Phi(B) = 1 - \phi_{1 \times 12} B^{12}$ and $\Theta(B) = 1 - \theta_1 B$. Their adequacy are further verified through the white noise test, where all p for autocorrelation within 24 lags were > 0.05 (Table 2). To compare with HMM using 1st order Fourier terms for seasonality, an AR(1) model was fitted with similar Fourier terms. The white noise test suggests it was insufficient to control for the seasonal pattern in the data, and therefore inadequate.

Fit statistics were further compared for the ARMA models (Table 3). Log-likelihood, AIC and BIC were all substantially improved in the two adequate ARMA models compared with the segmented regression model. $AR(1) \times (1)_{12}$ and $ARMA(0,1) \times (1,0)_{12}$ were, however, nearly identical in their fit. Their residual ACF and PACF graphs were also very similar, suggesting that the data could be adequately described by either model. Therefore, both were chosen to represent ARMA.

3.2 HMM

A multitude of HMM-Gaussian models without parameter constraints were fit. In this application, there are no theoretical guidance regarding the choice of the number of states (m) that can be assumed by the latent chain. Therefore, m was chosen by AIC. Similar to in the ARMA approach, stepwise model specification was used for the model of the dependent distributions. Fit statistics are displayed in Table 4.

Similar to ARMA, the fit statistics were substantially improved by adding seasonal and autoregressive features to the state-dependent distributions beyond the segmented model. The improvement in fit by incorporating additional autocorrelation to the basic HMM-Gaussian (as shown in Figure 2b) was relatively moderate. Overall, models with $m = 3$ performed best. In this case, adding the AR(1) term lowered AIC but increased BIC. Based on lowest AIC, the extended HMM with AR(1) state-dependent process was chosen to represent HMM.

The estimation of HMM involves some randomness. The experiment was run on batches of 50 trials to derive an average of the prediction SSE and an overview of the goodness-of-fit. Figure exhibits two typical distributions of the normal pseudo-residuals found in these trials. The upper panel with good fit is generated when the model was fitted with seed=50; the lower was fitted with seed=25. In the lower panel, the pseudo-residuals show skewness to the right and a long tail on the left. Overall, the models generated close-to-normal distributions in the normal pseudo-residuals, but a considerable proportion of the trials had one or more negative outlier.

Figure 7 represents the behavior of the latent Markov chain transitioning among 3 possible states. These states are all highly unstable, i.e. the chain almost never stays at the same states in two adjacent time periods.

3.3 Comparison on Prediction and Evaluation

The results presented so far have shown that both ARMA and HMM models were adequate for describing the data, and both were significant improvement over segmented regression. Although ARMA fitted the data perfectly with impeccable normal distribution in the standardized residuals, HMM performed substantially better in predicting future series, reducing the prediction SSE from 125,390 in the ARMA model to 108,271 (Table 5). HMM also fitted closer to the observed data (i.e. had smaller model residuals) compared with ARMA (Figure 8). For comparison with the extended HMM model, predictive accuracy was also evaluated for HMM-Gaussian model without the autoregression term; this model achieved lower accuracy than ARMA models (SSE= 142,349).

However, HMM had difficulty in evaluating the intervention effects of interest. Most trials to estimate the parameters with constraints have failed to generate equal estimates for the intervention effects independent of the state of the latent chain. The data reported in Table 5 were based on one successful trial, but its reliability is still questionable, according to warning messages from the computing software.

In contrast, both ARMA and segmented regression generated stable estimates. Estimates from the segmented regression model were noticeably different from the ARMA results. However, both methods concluded that the intervention had no significant effect on the outcome.

4. Discussions

In this exercise, time series analysis was explored with a special purpose of estimating the effects of a policy intervention. The problem was approached from two distinct traditions: one based on classical regression methods (ARIMA), another made fewer statistical assumptions and was more closely related to machine learning (HMM). By directly comparing their performance in different tasks in time series analysis, we found that ARIMA has been well-developed for intervention evaluation, but HMM outperforms ARIMA in prediction. Both approaches can accommodate features of univariate time series data.

4.1 Advantages of ARIMA

First of all, as a classical regression model, ARIMA is one of the routine methods in time series analysis and well-accommodated by standard statistical packages. Statisticians and social scientists have largely perfected its application for common tasks in social science research, such as intervention studies. This makes resources easier to find when solving problems similar to the one explored in this exercise.

Also due to its frequentist regression nature, it is relatively straightforward to draw inferences from ARIMA models in ways similar to linear regression. For example, although not used in this study, this feature can be useful for deriving confidence intervals for predictions.

Another advantage well-exemplified by this particular exercise is that the ARIMA model derived here was easy to interpret. In the $AR(1)_{12}$ process, observations in the same month of different years are positively correlated. This corresponds to annual patterns that occurs in many societal activities, including surgery. The other component of the model, $AR(1)$ or $MA(1)$, suggests that there could be natural fluctuation in the number of cancer surgery that counter-balances activities in the previous month. Although this interpretability may not be evident in more complicated series, this example shows that the feature can be useful in certain simple cases.

In comparison, the latent chain in the HMM model was difficult to interpret; the three states of the chain does not seem to correspond to any explanatory factors of cancer surgery. The instability of the states may be related to negative correlation between adjacent months, but not much more can be said about it.

4.2 Advantages of HMM

Despite its black-box nature on the Markov chain part, this example demonstrates HMM's utility even when the observed distribution is not an obvious state-dependent mixture. HMM was able to accommodate the structure of the data series. As mentioned earlier, a finite-state space, discrete-time homogeneous HMM has three sets of parameters – initial distribution, the $m \times m$ transition matrix, and parameters controlling the state-dependent distributions. This allows HMMs to have remarkable flexibility – each set of these parameters can be modeled with time-varying covariates, further adding to its versatility in handling time series data.

HMMs have more parameters than a similarly structured ARIMA model, owing to its Markov-chain related parameters and the multiplying effect of the number of states. Examination of Figure 8 may cause one to worry about over-fitting. In this example, however, HMM achieved better predictive accuracy on new data, suggesting HMM may have captured underlying structures of the

data series better than ARIMA. Unlike ARIMA, which models exclusively linear filters, HMM is better equipped to model nonlinear behavior.

In the HMM models, extending its basic form to incorporate direct autoregression in the observed series improved predictive accuracy substantially. This emphasizes the importance of the flexibility in HMM models.

HMM has unrealized potential in health research when prediction of future or past series is of interest. Although the confidence intervals of these predictions can be difficult to quantify theoretically, empirical methods such as bootstrapping may overcome this issue.

4.3 Future Work

This study did not solve the problem of using HMM to generate state-invariant estimates of the effect of an intervention. This issue may be solvable through optimization methods with constraints. It may also be possible to derive an overall expectation of the intervention effect as a probability-weighted sum of the state-dependent estimates.

Another area worth exploring is handling missing data to enable cross-validation. In this study, only the final year's observations were held out for model evaluation. This was because time series data is sequence-dependent; it is therefore less straightforward to hold out data in the middle of the series that later observations depend upon. This is however theoretically possible by estimating the model while treating the held-out observations as missing. However, a thorough understanding in estimating time series models with missing data is required.

5. Conclusions

In this example, segmented linear regression proved to be an inadequate model to examine intervention effects on a time series. The regression-based ARMA models and the machine-learning-oriented HMM models were both sufficient to describe a time series with trending, seasonal and autocorrelative patterns. While ARIMA was more well-developed to estimate intervention effects, HMM has better flexibility and may be more accurate in predicting future observations. HMM has unrealized potential in health policy research with time series data; current inconveniencies in its application to intervention studies is likely to be surmountable with existing statistical tools.

Bibliography

Bernal, James Lopez, Steven Cummins, and Antonio Gasparrini. 2016. "Interrupted Time Series Regression for the Evaluation of Public Health Interventions: A Tutorial." *International Journal of Epidemiology*, June, dyw098. doi:10.1093/ije/dyw098.

Harte, David. 2016. *HiddenMarkov: Hidden Markov Models* (version 1.8-7). https://cran.r-project.org/web/packages/HiddenMarkov/index.html.

Kontopantelis, Evangelos, Tim Doran, David A. Springate, Iain Buchan, and David Reeves. 2015. "Regression Based Quasi-Experimental Approach When Randomisation Is Not an Option: Interrupted Time Series Analysis." *BMJ* 350 (June): h2750. doi:10.1136/bmj.h2750.

Kutner, Michael, Christopher Nachtsheim, John Neter, and William Li. 2004. "Chapter 12 Autocorrelation in Time Series Data." In *Applied Linear Statistical Models*, 5th edition, 481–501. Boston: McGraw-Hill/Irwin.

McDowall, David, Richard McCleary, and Errol Meidinger. 1980. *Interrupted Time Series Analysis*. SAGE Publications.

Pickup, Mark. 2014. "Introduction to Time Series Analysis." In , 1 edition, 51–79. Los Angeles: SAGE Publications, Inc.

"R: ARIMA Modelling of Time Series." 2017. Accessed January 2. http://stat.ethz.ch/R-manual/R-devel/library/stats/html/arima.html.

Rob J Hyndman. 2014. "Forecasting Weekly Data." *Hyndsight*. March 4. http://robjhyndman.com/hyndsight/forecasting-weekly-data/.

SAS Institute Inc. 2014. "SAS/ETS® 13.2 User's Guide The ARIMA Procedure." https://support.sas.com/documentation/onlinedoc/ets/132/arima.pdf.

Visser, Ingmar, and Maarten Speekenbrink. 2016. "depmixS4: An R Package for Hidden Markov Models." https://cran.r-project.org/web/packages/depmixS4/vignettes/depmixS4.pdf.

Wagner, A. K., S. B. Soumerai, F. Zhang, and D. Ross-Degnan. 2002. "Segmented Regression Analysis of Interrupted Time Series Studies in Medication Use Research." *Journal of Clinical Pharmacy and Therapeutics* 27 (4): 299–309. doi:10.1046/j.1365-2710.2002.00430.x.

Zucchini, Walter, and Iain L. MacDonald. 2009. *Hidden Markov Models for Time Series: An Introduction Using R*. 1st edition. Boca Raton: Chapman and Hall/CRC.

Appendix: Tables and Figures from the Experiment

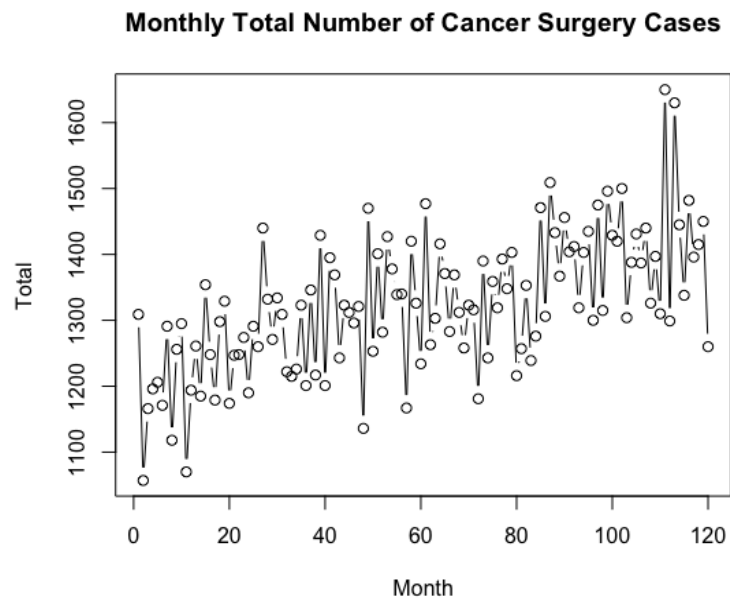**Monthly Total Number of Cancer Surgery Cases**

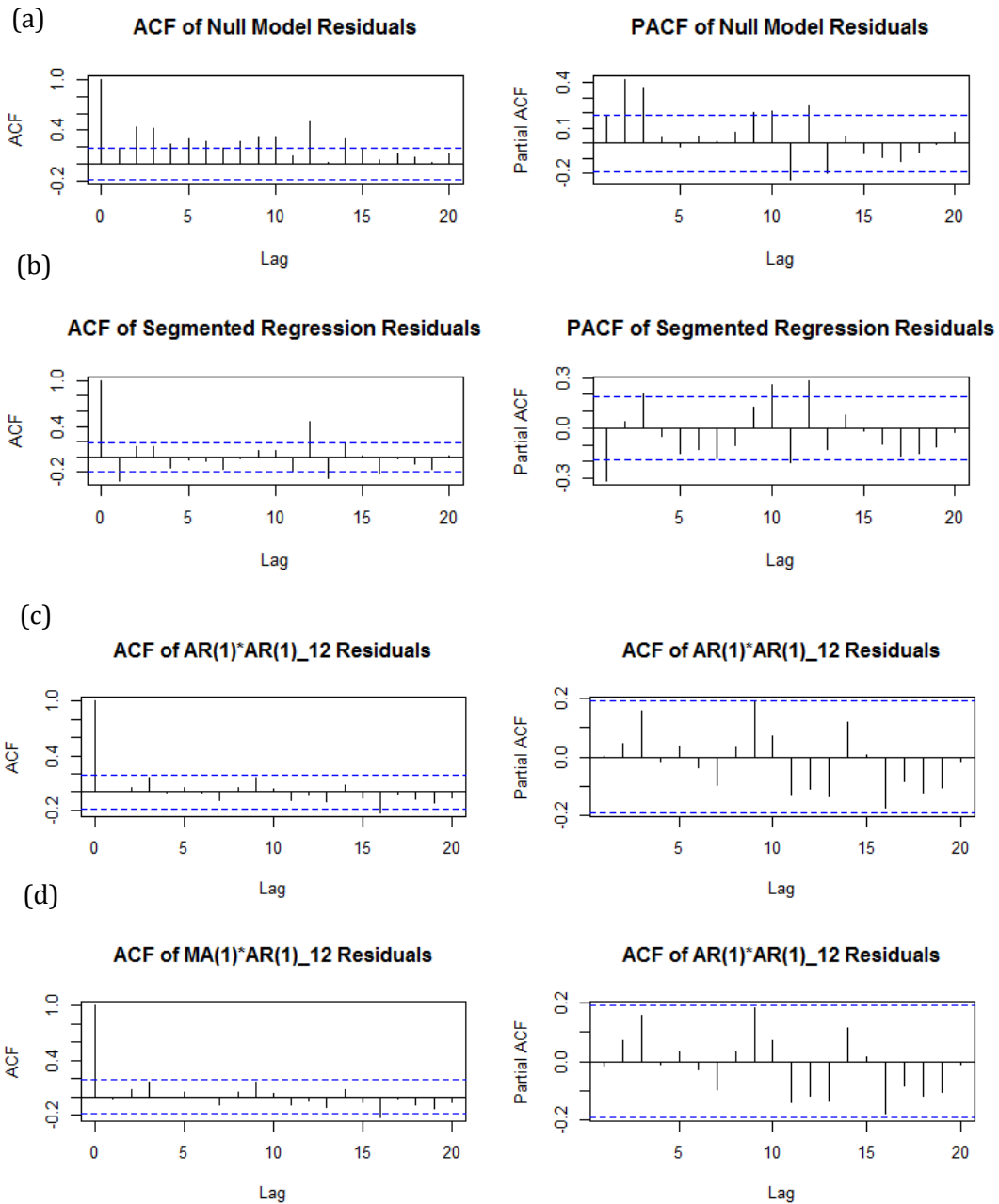Figure 3. Monthly Total Number of Inpatient Cancer Surgeries Performed in New York State, 1997-2006.

Figure 4. Autocorrelation and Partial Autocorrelation Functions of Regression Model Residuals.
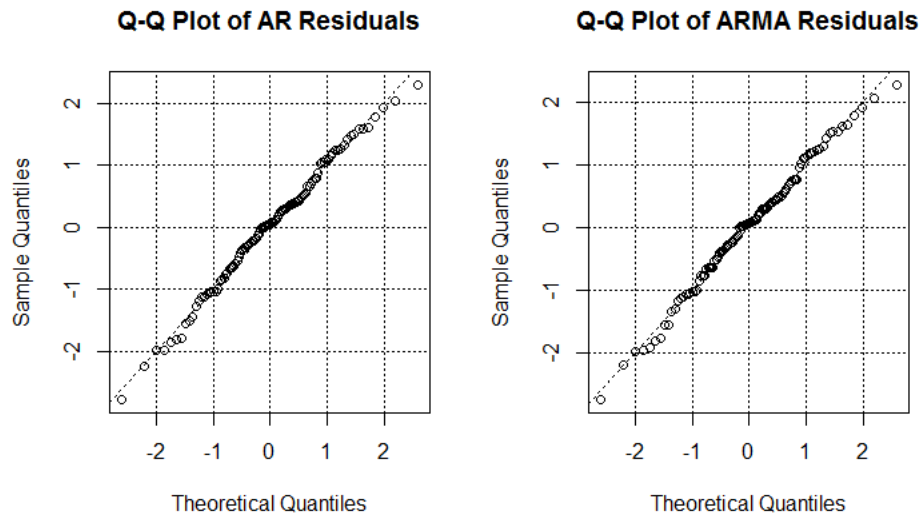
Figure 5. Q-Q plots of residuals from the ARMA models.

Table 2. Results of Portmanteau White Noise Tests for Regression Residuals in ARIMA Model Building.

Model	Lag = 6	Lag = 12	Lag = 18	Lag = 24
Null Model	2.63E-13	< 2E-16	< 2E-16	< 2E-16
Segmented Regression	0.007	2.23E-07	4.68E-09	6.00E-10
AR(1) with 1st Order Fourier Term for Seasonality	0.384	0.024	0.002	0.005
$AR(1)x(1)_{12}$	0.769	0.678	0.358	0.202
$ARMA(0,1)x(1,0)_{12}$	0.724	0.692	0.371	0.224

Table 3. Model Fit Statistics and from Select Regression Models

Model	Adequacy	Residual df	Log Likelihood	AIC	BIC
Null Model	No	107	-646.16	1296.32	1301.68
Segmented	No	104	-621.31	1252.63	1266.04
AR(1) with Fourier	No	101	-603.64	1223.28	1244.73
$AR(1)x(1)_{12}$	Yes	102	-603.49	**1220.97**	1239.75
$ARMA(0,1)x(1,0)_{12}$	Yes	102	-603.69	**1221.37**	1240.15

Table 4. Model Fit Statistics from HMM-Gaussian Models

	Model	Model DF	Log Likelihood	AIC	BIC
2-state Models	Null Model	7	-618.21	1250.42	1269.07
	Segmented Regression	13	-601.83	1229.65	1264.28
	+ 1st Order Fourier Seasonal Terms	17	-574.88	1183.75	1229.03
	+ 1st Order Autoregression	19	-569.84	1177.68	1228.29
	+ 2nd Order Autoregression	21	-569.19	1180.37	1236.31
3-state Models	Null Model	14	-615.90	1259.8	1297.08
	Segmented Regression	23	-593.00	1232.01	1293.27
	+ 1st Order Fourier Seasonal Terms	29	-555.65	1169.3	1246.54
	+ 1st Order Autoregression	32	-550.22	**1164.44**	1249.67
	+ 2nd Order Autoregression	35	-553.72	1177.44	1270.66
4-state Models	Null Model	23	-612.98	1271.95	1333.21
	Segmented Regression	35	-581.77	1233.54	1326.76
	+ 1st Order Fourier Seasonal Terms	43	-553.86	1193.72	1308.24
	+ 1st Order Autoregression	47	-548.91	1191.82	1317
	+ 2nd Order Autoregression	51	-534.86	1171.71	1307.55

HMM-Gaussian: Pseudo ResidualsMM-Gaussian: Q-Q Plot of Pseudo Resi
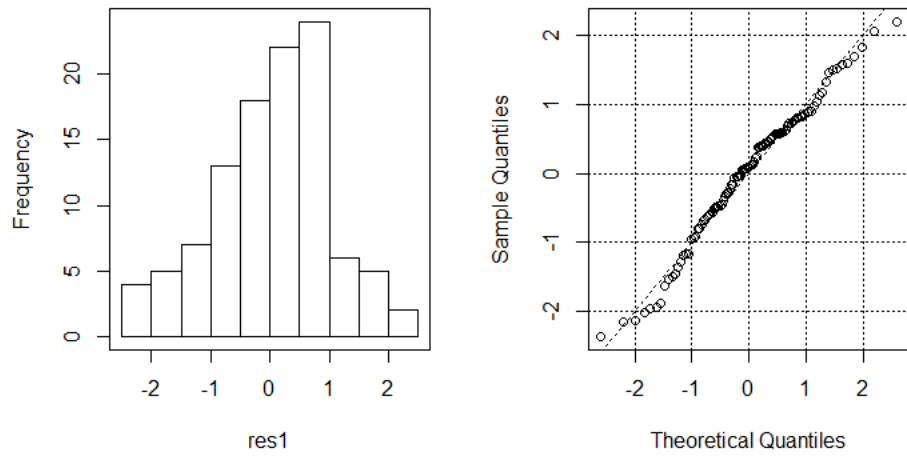
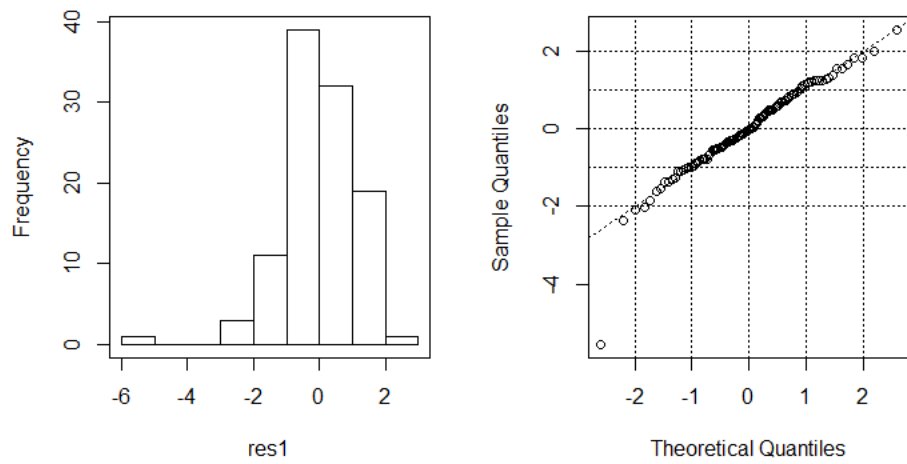**HMM-Gaussian: Pseudo ResidualsMM-Gaussian: Q-Q Plot of Pseudo Resi**

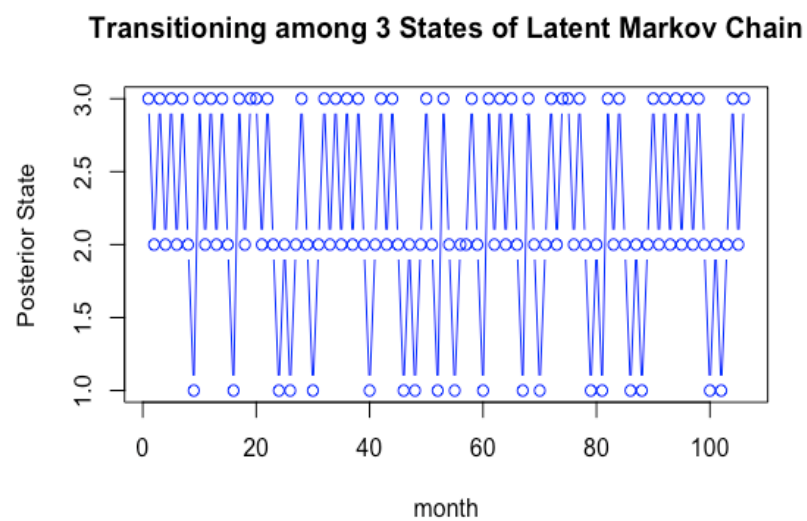Figure 6. Pseudo-residuals of HMM model with AR(1) state-dependent Gaussian process.

Figure 7. Transitioning Among States in the Latent Markov Chain.

Table 5. Predictive Accuracy and Estimates of Intervention Effects from Models

Model	Prediction SSE	Δ in intercept (SE)	Δ in slope (SE)
Segmented Regression	167141.5	-34.41 (30.0)	-0.51 (0.97)
ARMA(1,0)x(1,0)12	125742.6	-23.32 (20.7)	-0.85 (0.87)
ARMA(0,1)x(1,0)12	125390.5	-23.45 (20.7)	-0.85 (0.87)
HMM-Gussian	142348.7	N.A.	N.A.
HMM-Gaussian AR(1)	108270.8§	2.30¶	3.46¶

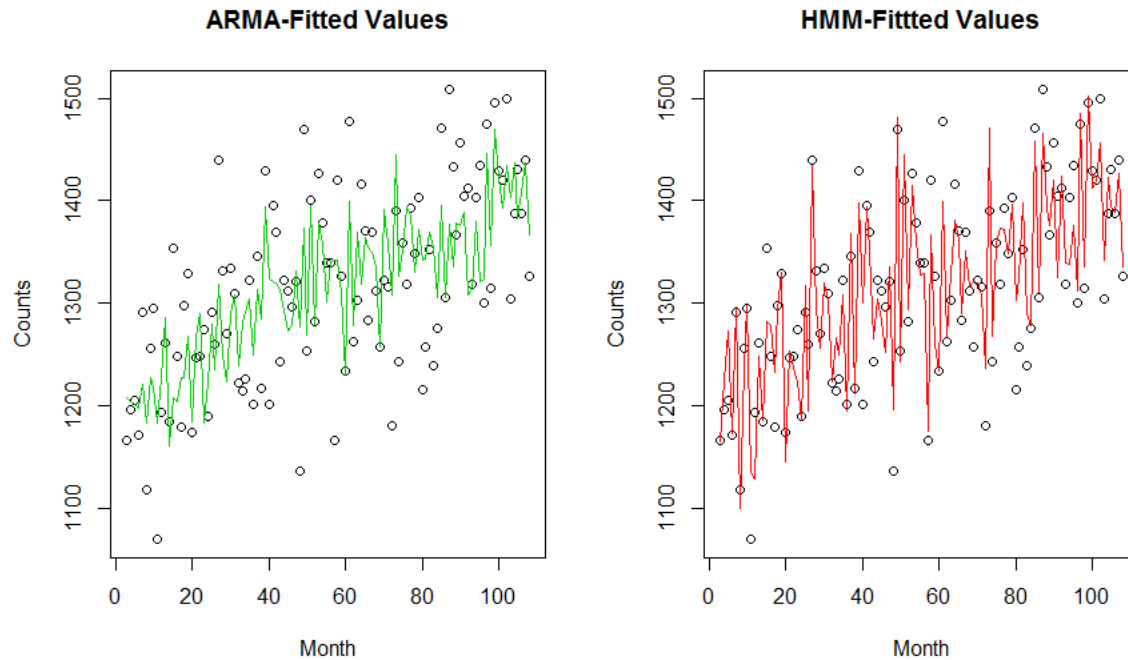§Based on 50 calculations with seeds 1 to 50. ¶Based on one successful run. May not be reliable.

Figure 8. Observed and Fitted Values from ARMA and HMM Models.