FREE ACADEMIC PAPER
SUBMITTED IN FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE
BACHELOR OF SCIENCE (B.Sc.)

Forecasting Emergency Patient Arrival Counts

Submitted to the Chair for Statistics and Econometrics II (520K)
(Professor Dr. Robert Jung)

Oliver Buchholz

Primary Reviewer: Prof. Dr. Robert Jung
Secondary Reviewer: Prof. Dr. Katja Schimmelpfeng

Stuttgart-Hohenheim, 24.11.2015

Contents

List of Symbols

(in their order of appearance)

ARIMA	Autoregressive Integrated Moving Average
SARIMA	Seasonal Autoregressive Integrated Moving Average
GLM	Generalised Linear Model
y	dependent variable, regressand
x	independent/exogenous variable, regressor
Y	random variable
$\mathbb{N}$	set of integers
$\cup$	set union
μ	intensity parameter of the Poisson distribution
$\mathbb{R}^+$	set of positive real numbers
$P(Y = y)$	probability function for Y taking on the value y
$\mathbb{E}[\cdot]$	expectation operator
$\sum_{t=1}^{T}$	sum running over the elements $t = 1, \ldots, T$
$Var(\cdot)$	variance operator
α	dispersion parameter of the negative binomial distribution
	level of significance
λ	intensity parameter of the negative binomial distribution
$\Gamma(\cdot)$	gamma function
t	time index
$\{Y_t\}_{t=1}^{T}$	stochastic process of random variable Y running over $t = 1, \ldots, T$
$\{y_t\}_{t=1}^{T}$	time series of observations y running over $t = 1, \ldots, T$
$\mathbf{x}_t$	$k \times 1$ vector of regressors for period t
x_{kt}	k-th element of $\mathbf{x}_t$
$\mathbf{y}_{t-q}$	$a \times 1$ vector of lagged dependent variables q periods apart from t
y_{at}	a-th element of $\mathbf{y}_{t-q}$

$\boldsymbol{\rho}_q$	$a \times 1$ vector of autocorrelation coefficients
ρ_{aq}	a-th element of $\boldsymbol{\rho}_q$
$f(y_t \mid \mathbf{x}_t, \mathbf{y}_{t-q}; \boldsymbol{\theta})$	conditonal density or probability mass function
$\boldsymbol{\theta}$	vector of parameters of $f(\cdot)$
$\boldsymbol{\beta}$	$k \times 1$ vector of regression coefficients
β	k-th element of $\boldsymbol{\beta}$
MLE	maximum likelihood estimator
QMLE	quasi-maximum likelihood estimator
$\prod_{t=1}^{T}$	product of elements running over $t = 1, \ldots, T$
$L_T(\cdot)$	average log-likelihood function
$\hat{\boldsymbol{\theta}}_{ML}$	MLE of $\boldsymbol{\theta}$
$\boldsymbol{\theta}_0$	vector of parameters of the true data generating process
$V_{ML}[\cdot]$	variance-covariance matrix of the MLE
$\xrightarrow{p}$	convergence in probability
$\hat{\boldsymbol{\theta}}_{QML}$	QMLE of $\boldsymbol{\theta}$
$V_{QML}[\cdot]$	variance-covariance matrix of the QMLE
$\hat{\beta}_k$	k-th estimated element of $\boldsymbol{\beta}$
$se(\cdot)$	standard error
z_t	Pearson residual for period t
ω_t	variance of the dependent variable, $Var(y_t)$
H_0	null hypothesis
T_{LB}	test statistic for the Ljung-Box portmanteau test
$\chi^2(df)$	χ^2-distribution
df	degrees of freedom
P	Pearson statistic
n	number of observations
AIC	Akaike information criterion
BIC	Bayesian information criterion
$f_{t,h}$	h-step ahead point forecast

y_{t+h}	realisation of the stochastiv process for period $t+h$
I_t	information set
$e_{t,h}$	h-step ahead forecast error
MSE	Mean Squared Error
$RMSE$	Root Mean Squared Error
$L(e_{t,h})$	loss function depending on the forecast error
$f^*_{t,h}$	optimal h-step ahead point forecast
$\int\limits_0^\infty$	integral from 0 to positive infinity
$\mu_{t+h\|t}$	mean for period $t+h$ conditional on I_t
$f_{t,h}(Y_{t+h}\|I_t)$	h-step ahead density forecast
$MISE$	Mean Integrated Squared Error
z	argument of the density functions used in the definition of $MISE$
$\pi(z)$	weight function of the $MISE$, set to one
ASE	Average Squared Error
P1-4	emergency arrivals of priority groups 1-4
ACF	autocorrelation function
$TEMP$	daily average air temperature (°C)
$TEMPMAX$	daily maximum air temperature (°C)
$TEMPMIN$	daily minimum air temperature (°C)
$RAIN$	rainfall (mm)
$SNOW$	snowfall (cm)
HUM	relative humidity (%)
$VAPRESS$	vapour pressure (hPa)
$D_{TEMPMAXi}$	variable = 1 for $TEMPMAX \geq i,\ i = 0$, 10, 15, 20, 30 °C, 0 otherwise
D_{RAINi}	variable = 1 for $RAIN > i,\ i = 1,\ 15$ mm, 0 otherwise
D_{SNOWi}	variable = 1 for $SNOW > i,\ i = 1,\ 5,\ 10$ cm, 0 otherwise

$D_{VAPRESSi}$	variable = 1 for $VAPRESS \geq i,\ i = 10,\ 15,\ 20$ hPa, 0 otherwise
$EA_t^{(i)}$	emergency arrivals for priority $i = 1, 2, 3, 4$ at time $t = 1, \ldots, 577$
$EA_{t-q}^{(i)}$	emergency arrivals for priority $i = 1, 2, 3, 4$ q periods before t
D_{iDAY}	variable = 1 for $i =$ MON, TUE, WED, THU, FRI, SAT, SUN, 0 otherwise
$D_{(af)PH}$	variable = 1 for day being public holiday (day after public holiday), 0 otherwise
D_{VAC}	variable = 1 for day being part of school holidays, 0 otherwise
D_{MONTH}	variable = 1 for MONTH = JAN, FEB, MAR, APR, MAY, JUN, JUL, AUG, SEP, OCT, NOV, DEC, 0 otherwise
D_{SUMMER}	variable = 1 for MONTH = JUN, JUL, AUG, 0 otherwise
D_{WINTER}	variable = 1 for MONTH = OCT, NOV, DEC, 0 otherwise

List of Tables

List of Figures

1 Introduction

For decades, the greatest part of time series econometrics and forecasting has been concerned with the analysis of economic time series such as the interest rate or the gross domestic product (González-Rivera 2013, pp. 2, for a historical sketch of time series analysis see Kirchgässner et al., Ch. 1.1). Starting by the end of the last century, however, the notion of something being "economic" changed or - more precisely - widened its scope. Nowadays, the economic realm not only embraces economic entities in a narrow sense, but also other parts of the society, such as science, education and the health care system (Graupe 2012, p. 639). Obviously, this tendency also widens the field for applications of econometric time series analysis, because more time series data is considered as being of economic relevance.

One example is the application of econometric methods to health care issues due to an increasingly competitive environment, especially with regard to hospitals. According to the Association of German Hospitals, hospitals face serious problems in refinancing their continuously increasing costs . An important cost driver in this context are personnel costs which tend to increase both due to tariff agreements and rising numbers of patients (Gesellschaft Deutscher Krankenhaustag 2012). In order to limit costs, it is therefore advantageous for hospitals to implement an efficient and effective mechanism for staff scheduling and rostering. Since personnel planning naturally depends on the expected demand for a hospital's health care services in the future, more efficient planning can be realised by using more sophisticated methods to predict the future demand. This is where econometric and especially time series methods come into play, because they are able "to predict a *future* event with some degree of *accuracy*" (González-Rivera 2013, p. 3) which is exactly what hospitals require to improve their planning processes.

Different attempts on this subject matter have been made in the past, involving different data and methods. However, one rather pronounced focus of the literature has been on predicting the number of future patients arriving in the emergency department of a hospital. This is reasonable, because compared to other patient groups, emergency patients can be regarded as arriving primarily stochastic and therefore as being cumbersome to include in any planning procedure, whether it is at a micro level for staff rosters, or at a macro level for budgeting (Sun et al. 2009, p. 1). In most of the cases, a so-called *Box-Jenkins approach* based on the *Autoregressive Integrated Moving Average* (ARIMA) class of time series models has been implemented to model and forecast emergency patient arrivals (see chapter 2 for details). Despite the fact, that these investigations identify certain key variables which influence the number of emergency patient arrivals - day of the week, month of the year or season - they disregard the pivotal characteristic of the data they use for forecasting: The number of emergency patient arrivals is necessarily a non-negative integer value, a fact that is not captured by ARIMA models. Additionally, the literature is far from being unanimous about all factors which affect the number of emergency arrivals and it is stated that there are important drivers which are only valid on a local level (Sun et al. 2009, p. 2).

The aim of this paper therefore is twofold: First, to apply a global theoretical framework that takes into account the time series as well as the count data character of the data at hand, in order to forecast emergency patient arrival counts. Second, it tries to verify the existence of predictors for future emergency arrivals already found by the literature and to identify additional predictors whose importance might be restricted to the local level under consideration. The paper seeks to achieve its aims by first reviewing the existing literature on the topic, followed by a presentation of the relevant econometric tools. Albeit being an investigation in applied econometrics, a special emphasis is made on the statistics underlying the modelling and forecasting process, because methods for count data "are not yet entirely standard in the econometric literature" (Winkelmann 2008, p. 8). Finally, the theory is applied to a dataset of daily data from an emergency department of a hospital in Brandenburg, Germany, ranging from 1st January 2013 to 30th September 2013. Patients in the dataset are stratified into four different priority groups, according to the time that is allowed to elapse before medical treatment starts. The entire modelling and forecasting process is separately applied to each group to reveal possible differences and to provide more information to decision makers in the hospital. The modelling and forecasting process is realised by following the three-step procedure presented in González-Rivera (2013, p. 202), where first the data is analysed, subsequently, time series models for counts - the Poisson and the Negbin II model - are estimated, evaluated and selected and finally, density forecasts are constructed to provide as much information as possible. The humble contribution this paper tries to make to the literature is the provision of density forecasts for future emergency arrivals, obtained from count data models. These can be especially useful for practitioners and their planning purposes, because - as Joseph A. Schumpeter put it once - "[t]hey are, by instinct, econometricians all of them" (Schumpeter 1933, p. 12).

The remainder of this paper is organised as follows: In chapter 2, already existing literature on the topic is reviewed and in chapter 3, the relevant statistical tools are presented, starting with discrete probability distributions and discussing related time series regression models and their estimation subsequently. In chapter 4, the statistical tools are applied to the dataset already mentioned, in order to forecast future emergency patient arrivals. Chapter 5 critically summarises this paper's contribution and gives an outlook on prospective research questions.

2 Related Literature

The literature of econometric applications to health care issues is widespread, both concerning the specific topics covered and methodological approaches used for investigation. The latter especially applies to the field of analysing and forecasting the number of patients arriving at a health care facility, be it a hospital or a doctor. Starting with pure linear regression analysis and time series regression, a methodologically rich literature has evolved, including Autoregressive Integrated Moving Average, seasonal Autoregressive Integrated Moving Average (SARIMA) and different types of count data models. Although the focus of this paper is on the application of count data models, the results of other investigations can deliver valuable insight into the mechanisms underlying the arrival of patients at a health care facility.

The article by Tandberg and Qualls (1994) who are - among others - using an ARIMA modelling strategy, can be considered as one of the first on the subject matter, trying to forecast the emergency patient volume. They find that "time series analysis can provide powerful quantitative short-range forecasts of future ED volume" (Tandberg and Qualls 1994, p. 305) with remarkably parsimonious models. Another finding which has been acknowledged in the literature ever since is the fact, that there is a daily and weekly pattern in the data of patient volume and arrivals (ibid.). Batal et al. (2001, p. 50) who forecast the daily patient volume in an emergency department using a linear regression approach specify this finding by the observation of a peak on Mondays followed by a stepwise decrease towards the end of the week and identify the day of the week as the strongest predictor for the daily patient volume. They also identify a seasonal pattern with a peak during the winter months and the lowest patient volume from April to August, being however not as predictive as the day of the week (ibid.). Aditionally, they state that the inclusion of the climatic variables minimum, maximum and average temperature as well as amount of precipitation does not improve the model's predictive ability by a considerable amount (Batal et al. 2001, pp. 48). This result contrasts with the work conducted by Jones et al. (2002, p. 299) who forecast the number of beds occupied due to emergency admissions using a SARIMA model. Apart from weekly and yearly seasonality, they also examine the effect of climatic factors on the number of emergency admissions, finding statistically significant negative correlation between the daily average temperature and emergency admissions (Jones et al. 2002, p. 300). This example shows, that while there is a broad agreement on seasonal effects and even their direction in the data of emergency patient arrivals, the existence of climatic effects remains controversial. Kam et al. (2010, p. 162), for instance, find a strong positive effect of the daily average temperature on the number of emergency patient arrivals, whereas Marcilio et al. (2013, p. 769) observe less ability to forecast when including temperature-related variables in their models. However, this stems at least partly from the fact that the data for each investigation is collected from a specific hospital in a specific geographical and climatic environment: While the data analysed by Batal et al. (2001, p. 48) is collected from a hospital in Denver, USA, Jones et al. (2002) use data from hospitals spread over the entire United Kingdom, Kam et al. (2010, p. 158) from a Korean hospital and Marcilio et al. (2013, p. 769) from Sao Paulo, Brazil.

Another approach with regard to the data under consideration is followed by Sun et al. (2009): They also use an ARIMA model but observations from the entire dataset of emergency patient arrivals are grouped into three acuity levels which are then analysed separately, which in fact yields diverging results. While the time series of patients belonging to the lowest acuity level revealed the usual weekly and yearly seasonality and was also predicted by a specific day being public holiday or not, the time series of patients belonging to the middle acuity level did not show any yearly seasonality (Sun et al. 2009, p. 1). Most strikingly, the time series of patients belonging to the highest acuity level did not show any seasonality and could not be predicted by any of the factors identified by the literature so far, which makes it especially difficult to predict future outcomes of the series (ibid.). Hence, it can be inferred that the inherent heterogeneity in an overall time series of patients arriving at a hospital's emergency department should be taken into account, in order to obtain meaningful results.

Besides factors driving the number of emergency patient arrivals which differ across publications, all studies cited so far share a statistical assumption underlying their analysis, which is only stated explicitly in Batal et al. (2001, p. 50): "On any given day the patient volume shows a normal distribution". In other words, the number of emergency patient arrivals is either observed or assumed to be normally distributed, which is why all studies mentioned above base their analysis on Gaussian models such as ARIMA or linear regression. Since a time series of emergency patient arrivals is a series of non-negative and integer numbers, this might not necessarily be feasible. As a consequence, many studies apply count data models to the respective time series. Marcilio et al. (2013, p. 769) use a Poisson General Linear Model (GLM) along with a SARIMA model, resulting in better ability to forecast with the former. McCarthy et al. (2008) use a Poisson regression model to predict the hourly number of emergency patient arrivals and observe the usual predictors day of week, season and day being a public holiday, as well as temperature and the amount of precipitation. Additionally, patients are stratified into a high and a low acuity level (McCarthy et al. 2008, p. 339). From a purely statistical point of view, the authors argue that a "Poisson model is plausible" (McCarthy et al. 2008, p. 338), because they observe the data following a positively skewed distribution and scrutinise the notion of prior patient arrivals predicting future patient arrivals inherent in time series models such as ARIMA (ibid.). For that purpose, they check the autocorrelation in the data, finding a certain degree of first-lag autocorrelation in the raw time series and no autocorrelation in the regression residuals when controlling for the factors mentioned earlier (McCarthy et al. 2008, p. 341). Finally, they also add lagged dependent variables to their model whose individual and joint effect is not statistically different from zero in a significant way, which in sum proves their hypothesis of arrivals being independent from each other to be true (ibid.). This result contrasts with other work in the area of count data models and their applications to health care topics, examples being Davis et al. (2003) and Jung et al. (2006) who analyse data of asthma presentations at a hospital in the Sydney metropolitan area, and Cardinal et al. (1999) who analyse data of cases of an infectious disease in the Montreal-Centre region, Canada. All of them use time series models for counts which are more sophisticated than the standard Poisson regression approach and they explicitly take into account the "rather pronounced dependence structure" (Jung et al. 2006, p. 2350) in the data. For that purpose, either *parameter-* or *observation-driven* models have been developed, where in the latter dependence is modelled by including lagged values of the dependent variable, while the former rely on a conditional mean function governed by a latent dynamic process (ibid.). Note that either way observations are not expected to be independent from each other, although the models are based on the assumption of Poisson-distributed counts.

To sum up the results already found by the literature, the following can be said: With regard to the crucial factors for the number of emergency patient arrivals, the day of the week - peak on Mondays, followed by a decrease towards the weekend -, the type of the day - public holiday or not - and - less pronounced - the month or season, respectively, are unquestioned. This is not the case for climatic factors such as the temperature and the amount of precipitation, where contradictory results depending on the area under consideration were found. Further heterogeneity was found in stratifying emergency patients into different acuity levels. Another aspect is the application of count data models instead of Gaussian models, which can result in a better ability to predict future arrival numbers. Still controversial remains the notion of arrivals being independent from each other which heavily affects both econometric modelling and forecasting and should therefore be tested in advance.

3 Theoretical Considerations

The present paper makes the attempt to forecast future arrivals of patients at the emergency department of a hospital. Hence, it is dealing with the number of a certain event at a specific point in time. The definition of this event - „number of emergency patient arrivals“ - implies, that the paper is in particular concerned with non-negative integers which are also known as counts. As pointed out in the second chapter, various studies use time series regression or ARIMA models to approach this subject matter. However, these models may in fact not be appropriate for the analysis of count data (Winkelmann 2008, p. 1).

3.1 Modelling Time Series Count Data

In general, the primary aim of econometrics is to model and estimate economic relationships between some dependent variable or regressand y and one or more independent variables or regressors x using statistical methods (Wooldridge 2013, p. 1). Having accomplished this first step, the model can serve to predict future outcomes of the dependent variable, which is called model-based forecasting (González-Rivera 2013, p. 13). Models like linear time series regression and ARIMA rely on the assumption of a dependent variable which follows a normal distribution conditional on the regressors (Wooldridge 2013, p. 113). Yet this is no suitable assumption for the analysis of count data: First, because the normal distribution belongs to the family of continuous probability distributions, whereas counts are of discrete nature (Georgii 2015, p. 52). Second, because the normal distribution is centered around its mean which one would not expect for the number of emergency patient arrivals. To avoid these shortcomings, probability models which take into account the specific nature of the data will be presented below.

3.1.1 Probability Models

As set out by Cameron and Trivedi (1986, p. 3), the standard model for count data is the Poisson distribution. Since the same statement applies to their monograph, being the standard textbook for count data, the following lines are based on Cameron and Trivedi (1986, pp. 3) as well as Winkelmann (2008, Chapter 2.2).

Consider Y being a discrete random variable defined over $\mathbb{N} \cup 0 = \{0, 1, 2, \ldots\}$, which means that Y only takes non-negative and integer values. Y follows a *Poisson distribution* with intensity parameter $\mu \in \mathbb{R}^+$ denoted by $Y \sim Po(\mu)$ if and only if the *probability mass function* is defined as

$$P(Y = y) = \frac{e^{-\mu}\mu^y}{y!}, \quad y = 0, 1, 2, \ldots \tag{1}$$

Note, that the expression in (1) is referred to as *probability mass function* rather than probability density function, which is its analogue for the continuous case. The probability mass function returns the probability of the random variable Y taking on the value y, so that y is a realisation of Y which obviously must also be a non-negative integer. Additionally, the probability mass function allows for the derivation of the distribution's expected value resulting in

$$\mathbb{E}(Y) = \sum_{y=0}^{\infty} \mu \frac{e^{-\mu}\mu^y}{y!} = \sum_{y=1}^{\infty} \mu \frac{e^{-\mu}\mu^y}{y!} = \mu \sum_{y=1}^{\infty} \frac{e^{-\mu}\mu^{y-1}}{(y-1)!} = \mu. \tag{2}$$

In another way which will not be deepened here, it is possible to obtain the Poisson distribution's variance as well, which is given by

$$Var(Y) = \mu \quad . \tag{3}$$

Remarkably, the Poisson distribution implies equality of mean and variance, also known as *equidispersion*. With regard to the already mentioned normal distribution which is centered around the mean, the Poisson distribution is characterised by positive skewness, resulting in a longer right tail of the probability mass function.

It is sensible to assume, that the Poisson distribution, given its characteristics, is better capable to fit the emergency patient arrival counts than a normal distribution. However, when modelling real-world data, the assumption of equidispersion can be too restrictive. Other situations, which can arise, are both underdispersion, $\mathbb{E}(Y) > Var(Y)$, and overdispersion, $\mathbb{E}(Y) < Var(Y)$. Since the Poisson distribution is a one-parameter distribution, it cannot capture these situations adequately.

Another frequently used probability distribution for count data which can serve as an alternative to the Poisson distribution is the *negative binomial distribution* (see Winkelmann 2008, pp. 20 for further discussion). The same random variable Y, defined above, follows a negative binomial distribution with parameters $\alpha \geq 0$ and $\lambda \geq 0$, denoted by $Y \sim Negbin(\alpha, \lambda)$ if the probability mass function is as follows:

$$P(Y = y) = \frac{\Gamma(\alpha + y)}{\Gamma(\alpha)\Gamma(y+1)} \left(\frac{1}{1+\lambda}\right)^{\alpha} \left(\frac{\lambda}{1+\lambda}\right)^{y}, \quad y = 0, 1, 2, \ldots \tag{4}$$

Interpretation of $P(Y = y)$ and y is completely analogous to expression (1), $\Gamma(\cdot)$ denotes the gamma function given by $\Gamma(s) = \int_0^{\infty} z^{s-1}e^{-z}dz$ for $s > 0$. Note, that in contrast to the Poisson distribution, the negative binomial distribution requires two parameters to be completely specified. This will result in more flexibility when it comes to the derivation of regression models, as will be pointed out in the subsequent part of this chapter. The mean and the variance of the negative binomial distribution are given by

$$\mathbb{E}(X) = \alpha\lambda \tag{5}$$

and

$$Var(X) = \alpha\lambda(1+\lambda) = \mathbb{E}(X)(1+\lambda). \tag{6}$$

Applying the definition $\lambda \geq 0$ to (6) shows that, in general, overdispersion is a common characteristic of the negative binomial distribution, because its variance exceeds its mean. Obviously, this overdispersion tends to zero as $\lambda \to 0$ (Winkelmann 2008, pp. 20). Just as the Poisson distribution, the negative binomial distribution is positively skewed with a longer right tail of its probability mass function. Figure 1 depicts both of the distributions outlined above for $\mathbb{E}(Y) = 3.5$ and $Var(Y)/\mathbb{E}(Y) = 2$ in the case of the negative binomial distribution.

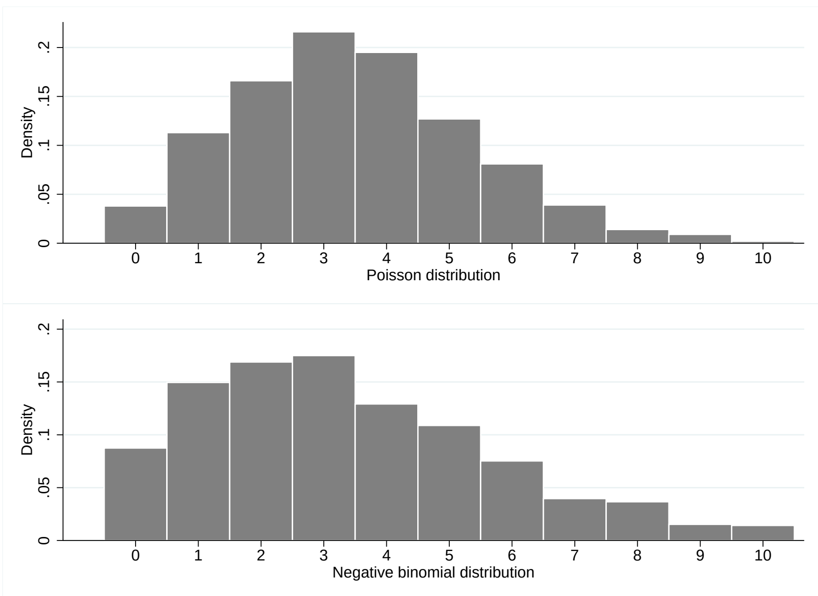

Figure 1: Count data distributions ($\mathbb{E}(Y) = 3.5$, following Winkelmann (2008, p. 29)

3.1.2 Derivation of Regression Models

To become tools for econometric analysis, it is necessary for the probability distributions to be integrated in a regression framework. Before doing so, it is reasonable to clarify the specific context of the present paper in more formal terms and to give some definitions. Since it is the aim of this investigation to model and forecast emergency patient arrival counts, it is concerned with constructing a model for realisations y of a random and dependent variable Y at a given point in time t, or more briefly, a time series model. A possibility to include the temporal dimension in the analysis of random variables is the concept of a *stochastic process*, denoted by $\{Y_t\}_{t=1}^{T} = \{Y_1, Y_2, \ldots, Y_T\}$ (González-Rivera 2013, pp. 54). The notation implies, that a stochastic process consists of a collection of random variables with a subindex indicating the time, running from 1 to T. Apart from the notation, the temporal dimension is the reason why each of the random variables in the stochastic process has one and only one realisation y_t; each period t is unique, each period's realisation is therefore irreversible. The sample of all realisations from the stochastic process forms a time series $\{y_t\}_{t=1}^{T} = \{y_1, y_2, \ldots, y_T\}$ which corresponds to observable data, whereas the stochastic process is unobservable. Model-based forecasting tries to exploit this time series data to infer the characteristics of the underlying stochastic process in order to be able to predict its future outcomes (ibid.). In a regression context, this means that the set of regressors $\mathbf{x}_t$ can be expanded by lagged values of the dependent variable from some prior period $t-q$, y_{t-q}, which are known as *autoregressive* elements. The correlation between y_t and some prior value y_{t-q} is therefore known as *autocorrelation* and for all values which are q periods apart from each other and belong to the same stochastic process, the *autocorrelation function* $\rho : q \to \rho_{Y_t, Y_{t-q}}$ is defined as

$$\rho_q = \frac{cov(Y_t, Y_{t-q})}{\sqrt{var(Y_t)}\sqrt{var(Y_{t-q})}}, \tag{7}$$

where ρ_q is the correlation coefficient for two realisations of the stochastic process $\{Y_t\}_{t=1}^{T}$ being q periods apart from each other (González-Rivera 2013, pp. 65). Note, that when constructing a model for the emergency patient arrival counts, the restriction $Y_t \in \mathbb{N} \cup 0, \ \forall t$ on the stochastic process has to be made for obvious reasons.

It is the task now to combine all results mentioned so far in a theoretical framework, that allows to model and forecast the time series for the emergency patient arrival counts. It therefore has to take explicitly into account that the dependent variable is a count and can be considered as the realisation of a stochastic process, which implies serial dependence across observations. Cameron and Trivedi (2013, Ch. 7) present various models for count data which are able to capture temporal dependency of the dependent variable in different ways. One straightforward way is to specify an autoregressive model, where, based on a basic count data regression model, lagged dependent variables are included as additional regressors, which is referred to as an observation-driven model, since it is specified by using past observations as predictors (Cameron and Trivedi 2013, pp. 267, 281). As a next step, this "augmented" regression model for counts is presented, first in the Poisson and subsequently in the Negbin II specification, following Cameron and Trivedi (2013, Ch. 2, 3 and 7) and Martin et al. (2013, pp. 339). Starting with the probability mass function of the Poisson distribution in (1), it may now be assumed, that the dependent variable y_t given the $k \times 1$ vector of qualitative or quantitative regressors $\mathbf{x}_t$ and lagged values $\mathbf{y}_{t-q}$ is Poisson distributed with probability mass function

$$f(y_t|\mathbf{x}_t, \mathbf{y}_{t-q}; \theta) = \frac{e^{-\mu_t}\mu_t^{y_t}}{y_t!}, \tag{8}$$

$$y_t = 0, 1, 2, \ldots; \quad k = 1, 2, \ldots, \ t-1,$$

where the set of parameters is given by $\theta = \{\mu_t\}, \ \mu_t \in \mathbb{R}^+, \forall t$, since the Poisson distribution is fully characterised by one parameter. Although it is a strong assumption to specify the entire distribution of the dependent variable explicitly, the gain is to obtain a positively skewed conditional probability mass function which might be appropriate for modelling problems such as emergency patient arrivals. By recognising that the Poisson distribution is a one parameter distribution, it becomes clear, that there is only one possibility to introduce a regression model, that is, to model the mean or intensity parameter μ_t of the distribution. The intention is, that explanatory variables influence the dependent variable through the intensity parameter (Winkelmann 2008, pp. 64), which is most commonly specified as

$$\mu_t = exp(\mathbf{x}_t'\boldsymbol{\beta} + \mathbf{y}_{t-q}'\,\boldsymbol{\rho}_q), \tag{9}$$

where for k being the number of regressors including a constant, $\boldsymbol{\beta}$ is a $k \times 1$ parameter vector and $\boldsymbol{\rho}_q$ is a $a \times 1$ vector containing autoregressive parameters of lag order q for a number of a lagged values of y included in the model. This implies a set of parameters $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{\rho}_q\}$. Since one condition for a suitable model of emergency patient arrivals is, that it produces non-negative outcomes, the exponential of the right-hand term in (9) is taken. From the properties of the Poisson distribution in chapter 3.1.1 it can be seen, that (9) also implies

$$\mathbb{E}[y_t|\mathbf{x}_t, \mathbf{y}_{t-q}] = \mu_t = exp(\mathbf{x}_t'\boldsymbol{\beta} + \mathbf{y}_{t-q}'\,\boldsymbol{\rho}_q), \tag{10}$$

which is known as an exponential or log-linear mean function, since the logarithm of (10) is linear in the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\rho}_q$. The interpretation of the parameters - or regression coefficients - will not be discussed in further detail here, because it is of no primary interest for the forecasting purpose. Nevertheless, some remarks can be found in Appendix A. Note, that Cameron and Trivedi (2013, p. 281) advise caution when including lagged values of the dependent variable in the mean function, because $\rho_q > 1$ implies $\rho_q y_{t-q} \geq 0$ which might result in an explosive model.

An additional remark has to be made with regard to the inherent notion of equidispersion in the Poisson distribution mentioned earlier. This notion implies, that in the Poisson regression model, the conditional mean equals the conditional variance, $\mathbb{E}[y_t|\mathbf{x}_t, \mathbf{y}_{t-q}] = Var(y_t|\mathbf{x}_t, \mathbf{y}_{t-q})$, a situation, which arises seldom in applied econometrics. The Poisson regression model might therefore be inappropriate and too restrictive when data is observed to be overdispersed. Cameron and Trivedi (2013, pp. 74) present an alternative way for model specification in cases where overdispersion is an issue. One possible way is to model the conditional variance as a function of the mean, which remains specified as in (9). The most common specification is known as Negbin II variance function, given by

$$Var(y_t|\mathbf{x}_t, \mathbf{y}_{t-q}) = \mu_t + \alpha^{-1}\mu_t^2, \tag{11}$$

where α is the *dispersion parameter* which needs to be estimated separately. The case where α is found to be statistically different from zero is considered as a test for overdispersion. This setting requires a different regression framework than the one presented in (8) through (10), allowing for two parameters, μ_t and α. In order to set up this framework, the expression in (5) is converted into a mean parametrisation $\alpha\lambda = \mu_t$ (Winkelmann 2008, p. 21). The dependent variable y_t is then assumed to be distributed negative binomially given the $k \times 1$ vector of regressors $\mathbf{x}_t$ and lagged values $\mathbf{y}_{t-q}$ with probability mass function

$$f(y_t|\mathbf{x}_t, \mathbf{y}_{t-q}; \boldsymbol{\theta}) = \frac{\Gamma(\alpha + y_t)}{\Gamma(\alpha)\Gamma(y_t + 1)} \left(\frac{\alpha}{\alpha + \mu_t}\right)^{\alpha} \left(\frac{\mu_t}{\alpha + \mu_t}\right)^{y_t} \tag{12}$$

$$t = 0, 1, \ldots, T; \; y = 0, 1, 2, \ldots,$$

where it holds now that $\boldsymbol{\theta} = \{\mu_t, \alpha\}$ and the other elements of the equation have the same interpretation as given in (4) and (9), respectively. This Negbin II regression model is more flexible when it comes to model overdispersed data, since the conditional variance must not necessarily equal the conditional mean. However, the regression models are obviously nested, because for $\alpha = 0$ the Negbin II model equals the Poisson model, which is why above derivation of the Negbin II model is referred to as *Poisson-gamma mixture* interpretation (for details see Cameron and Trivedi 2013, pp. 117). Since the conditional mean of the Negbin II model equals the specification given in (10), the same caution has to be given to the autoregressive coefficients being less than one.

3.2 Estimation and Diagnostics

Just as all econometric models do, the Poisson and Negbin II model refer to some unobserved *data generating process*, but rely in fact on a sample of observed data, in the case of this study on emergency patient arrival counts. Parameters which are used in the specification of the models therefore have to be estimated, first and foremost the $k \times 1$ and $a \times 1$ coefficient vectors $\boldsymbol{\beta}$ and $\boldsymbol{\rho}_q$ as well as the dispersion parameter α in the Negbin II model. Since the conditional mean of both models is specified as a non-linear function in (10), Ordinary Least Squares (OLS) estimation cannot be applied. The standard estimator is obtained by *maximum likelihood* (ML) estimation, which will be referred to as maximum likelihood estimator (MLE) from now on (Cameron and Trivedi 2013, pp. 22), another way to obtain parameter estimates is *quasi*-maximum likelihood (QML) estimation, resulting in the quasi-maximum likelihood estimator (QMLE) (Cameron and Trivedi 2013, pp. 72).

3.2.1 Maximum Likelihood Estimation

Since the present paper partly uses ML estimation to obtain parameter estimates for the model of emergency patient arrivals and heavily relies on the entire ML framework, some basic theory will be presented in the following paragraph. A comprehensive and very thorough presentation of ML theory along with examples involving models for counts is given in Martin et al. (2013). A necessary step to obtain the MLE of the model's parameters is to define the model's joint probability mass function which utilises all information available from $t = 0, 1, \ldots, T$ (Martin et al. 2013, pp. 9). In doing so, it is important to take into account the specific nature the data generating process is assumed to have. As a direct implication of (10), the data generating process is assumed to depend on its own lags and the time series of explanatory variables $\{\mathbf{x}_1, \ldots, \mathbf{x}_T\}$ for the present application. Following again Martin et al. (2013, pp. 11, eq. 1.7), the joint probability mass function for this situation is given by

$$f(y_1, \ldots, y_T | \mathbf{x}_1, \ldots, \mathbf{x}_T; \boldsymbol{\theta}) = \\ f(y_1|\mathbf{x}_1; \boldsymbol{\theta}) \prod_{t=2}^{T} f(y_t|y_{t-1}, \ldots, y_1, \mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, \mathbf{x}_1; \boldsymbol{\theta}), \qquad (13)$$

where $\boldsymbol{\theta} \in \mathbb{R}^q$ is a $q \times 1$ vector of parameters to be estimated, so that $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{\rho}_q\}$

for the Poisson and $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{\rho}_q, \alpha\}$ for the Negbin II model (Cameron and Trivedi 2013, pp. 23). Now, the intuition of ML estimation is the inverse interpretation of the joint probability mass function in (13): $f(\cdot)$ is a function of $\boldsymbol{\theta}$ for *given* y_t, because the time series $\{y_1, \ldots, y_T\}$ is regarded as being already realised and therefore non-stochastic. The task becomes then to estimate the value $\boldsymbol{\theta}$ "which is 'most likely' to have generated the observed data" (Martin et al. 2013, p. 12). For that purpose, the average *log-likelihood function* $\ln\, L_T(y_1, \ldots, y_T|\mathbf{x}_1, \ldots, \mathbf{x}_T; \boldsymbol{\theta})$ is generally defined as

$$\begin{aligned} \ln\, L_T(\boldsymbol{\theta}) = &\frac{1}{T}\, \ln\, f(y_1|\mathbf{x}_1; \boldsymbol{\theta}) \\ &+ \frac{1}{T} \sum_{t=2}^{T}\, \ln\, f(y_t|y_{t-1}, \ldots, y_1, \mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, \mathbf{x}_1; \boldsymbol{\theta}), \end{aligned} \tag{14}$$

where, for reasons of brevity, $\boldsymbol{\theta}$ now is the only argument (ibid.). Note, that for the modelling problem at hand, this expression cannot be simplified, since restrictive assumptions such as the dependent variable being independent and identically distributed are avoided. The MLE $\hat{\boldsymbol{\theta}}_{\text{ML}}$ of $\boldsymbol{\theta}_0$ is defined as the value maximising the average log-likelihood function in (14), $\hat{\boldsymbol{\theta}}_{\text{ML}} = \text{argmax}\ L_T(\boldsymbol{\theta})$, which in some cases can be achieved by standard calculus (ibid.). However, for the case of the Poisson and Negbin II model, the first-order conditions obtained by differentiating the average log-likelihood function once with respect to the unknown parameters, are non-linear and therefore provide no analytical solution to calculate parameter estimates (Cameron and Trivedi 2013, p. 23).

Another shortcoming of non-linearity is a lack of analytical results with regard to small sample properties of $\hat{\boldsymbol{\theta}}_{\text{ML}}$ (Winkelmann 2008, p. 80). Given the so-called *regularity conditions* (for details see Cameron and Trivedi 2013, pp. 25) hold, at least asymptotic properties can be derived. Namely, these properties are asymptotic efficiency and consistency, as well as approximate asymptotic normality with mean $\boldsymbol{\theta}_0$ and variance $\text{V}_{\text{ML}}[\hat{\boldsymbol{\theta}}_{\text{ML}}]$, $\hat{\boldsymbol{\theta}}_{\text{ML}} \overset{a}{\sim} N(\boldsymbol{\theta}_0, \text{V}_{\text{ML}}[\hat{\boldsymbol{\theta}}_{\text{ML}}])$, which are derived in Martin et al. (2013, Ch. 2) for the general case and in Cameron and Trivedi (2013, pp. 25) as well as Winkelmann (2008, pp. 80) for count data models. In each of these monographs it is stated along with the derivations, that the results require a correctly specified density - i.e. probability mass function - in the sense that asymptotic consistency means, that the MLE converges in probability to the parameter vector $\boldsymbol{\theta}_0$ of the data generating process, $\hat{\boldsymbol{\theta}}_{\text{ML}} \overset{p}{\to} \boldsymbol{\theta}_0$. This implies, that the observations $\{y_1, \ldots, y_T\}$ must necessarily follow a Poisson or negative binomial distribution defined in (8) and (12), respectively. It turns out, that correct specification of the entire joint probability mass function is the price which is to pay for the useful properties of the MLE.

3.2.2 Quasi-Maximum Likelihood Estimation

It is due to its own restrictiveness, that the assumption of a correctly specified probability mass function made in ML estimation is often violated in practice. For instance, in the Poisson regression model, one very common deviation from probabilistic assumptions which can be found in real data is the presence of overdispersion, with conditional variance exceeding the conditional mean. For this case, Cameron and Trivedi (2013, pp. 70) propose different ways to continue: Obviously, it is an option to work with a better parametric model, namely the Negbin II instead of the Poisson model, where it is possible to model overdispersion via the conditional variance explicitly. This procedure, however, is also carried out within the ML framework and requires correct specification of the joint negative binomial probability mass function. Another option, known as *quasi*-ML (QML) estimation, is to continue with the misspecified Poisson model within the ML framework which is then valid under less restrictive assumptions than the standard ML approach. Again, Martin et al. (2013, Ch. 9) provide a very thorough discussion of this issue, applications to count data models can be found in both Cameron and Trivedi (2013, pp. 70) and Winkelmann (2008, pp. 87).

Given, that the joint probability mass function of the model, as expressed in (13), is misspecified, the log-likelihood function likewise is. Maximising this function as described in the previous section leads to the QML estimator (QMLE), denoted as $\hat{\boldsymbol{\theta}}_{\text{QML}}$. Building upon work originally conducted by Gourieroux et al. (1984), Cameron and Trivedi (2013, pp. 31) point out, that, regardless of the true data generating process for the dependent variable, the QMLE is consistent, $\hat{\boldsymbol{\theta}}_{\text{QML}} \overset{p}{\to} \boldsymbol{\theta}_0$, as long as the assumed probability mass function belongs to a certain family of distributions and the conditional mean in (10) is correctly specified. Although justification and derivation of the assumptions lie well beyond the scope of this paper, the importance of the result for applied work has to be stressed: Using QML estimation leads to consistent parameter estimates without the requirement of complete and correct specification of the joint probability mass function, that is, the true data generating process must not follow a equidispersed Poisson distribution.

So far, it has been stated that the MLE of a correctly specified model approximately follows a normal distribution, if the sample is large enough. Albeit being a consistent estimator for $\boldsymbol{\theta}_0$, this statement is not necessarily true for the QMLE, which results in wrong inferences due to underestimated standard errors, if the ML variance matrix - which is based on the assumption of equidispersion - is used (Winkelmann 2008, pp. 91). Again, under the assumption of a correctly specified conditional mean function, the problem of inference within the QML framework can be solved by applying *robust* or *Eicker/Huber/White standard errors*[1]. In a nutshell, the aim is to adjust the "wrong" ML variance matrix for possible overdispersion, which can be achieved by deriving a consistent estimator for the conditional variance which, in turn, is the crucial part of the original ML variance matrix (for a more detailed treatment of robust estimation see Winkelmann 2008, pp. 91 and Cameron and Trivedi 2013, p. 73). Combining the consistently estimated conditional variance and the QMLE's consistency, this results in $\hat{\boldsymbol{\theta}}_{\mathrm{QML}}$ being approximately normal with mean $\boldsymbol{\theta}_0$ and adjusted variance matrix $\mathrm{V}_{\mathrm{QML}}[\hat{\boldsymbol{\theta}}_{\mathrm{QML}}]$, $\hat{\boldsymbol{\theta}}_{\mathrm{QML}} \overset{a}{\sim} N(\boldsymbol{\theta}_0, \mathrm{V}_{\mathrm{QML}}[\hat{\boldsymbol{\theta}}_{\mathrm{QML}}])$ (Cameron and Trivedi 2013, p. 73). Cameron and Trivedi (2013, p. 71) also state, that for most applications Poisson QML estimation as presented in this section, or Negbin II ML estimation as presented in the previous one, are sufficient in terms of flexibility to capture possible overdispersion. This is why the present study also relies on these estimation techniques and does not discuss further options.

3.2.3 Postestimation

Even though it is not the purpose of this paper to identify and explain causal relationships driving the number of emergency arrivals, but to forecast them as precisely as possible, some attention has to be drawn to statistical inferences made upon parameter estimates. This stems from the fact, that regression coefficients which do not differ from zero on a predetermined level of significance α, are dropped from the model which serves as a forecasting tool afterwards. This level of significance is chosen to be $\alpha = 0.05$ in the present study, testing for significance of single coefficients is performed using the usual t-test with test statistic $\hat{\beta}_k / se(\hat{\beta}_k)$, where $\hat{\beta}_k$ is an estimated regression coefficient from the $k \times 1$ vector $\hat{\boldsymbol{\beta}}$ and could also be replaced by an estimated autocorrelation coefficient from the vector $\hat{\boldsymbol{\rho}}_{t-q}$ (for a detailed presentation of the t-test see Wooldridge 2013, pp. 120). This postestimation procedure is completely analogue to the standard procedure within the classical linear regression model. Still, there are other issues which differ considerably and therefore have to be given particular attention, especially the analysis of regression residuals.

For a linear regression model, the regression residual is defined in a straightforward way as the difference between fitted and observed values (Cameron and Trivedi 2013, p. 178). However, several assumptions have to be made to obtain residuals having the desirable properties of being normally distributed with mean zero and constant variance one. Homoskedasticity is one of them, stating that the residual variance stays constant across all observations in the sample. This assumption is not fulfilled for the standard residuals - the so-called *raw residuals* - from the Poisson or Negbin II model, which are heteroskedastic even in large samples (Cameron and Trivedi 2013, p. 179). To correct for this feature of

[1] Named after mathematical statisticians Friedhelm Eicker, Peter J. Huber and econometrician Halbert J. White.

count data models, different residuals which make different adjustments to the raw residuals have been developed. In this study, the exclusive focus is on the *Pearson residual*, defined as

$$z_t = \frac{(y_t - \hat{\mu}_t)}{\sqrt{\hat{\omega}_t}}, \tag{15}$$

where y_t is the observation in period t, $\hat{\mu}_t$ is the estimated conditional mean specified according to equation (10) and $\hat{\omega}_t$ is an estimate of the variance ω_t of y_t, which means that the denominator consists of the standard error $se(\hat{\mu}_t)$ (Cameron and Trivedi 2013, p. 179). So, from above it follows that for the Poisson model $\omega_t = \mu_t$ and for the Negbin II model $\omega_t = \mu_t + \alpha^{-1}\mu_t^2$. Obviously, the standard error in the denominator is what distinguishes the Pearson from the raw residual and it serves as a correction for heteroskedasticity, in order to yield properties which come as close as possible to the desired symmetric distribution with zero mean and variance equal to one. And the Pearson residual indeed shows at least two of these properties, having zero mean and a constant variance of one in large samples, but being asymmetrically distributed at the same time (ibid.). When the residuals are obtained within the estimation step of the modelling procedure in chapter 4, they are checked for the two properties mentioned first. Additionally, they help to assess the model specification: Since it is one feature of a sophisticated forecasting model to capture a maximum of the dependence that can be seen in the time series of the data by inspection of the autocorrelation function, no dependence should be left in the regression residuals after fitting the model. More technically, the Pearson residuals will be uncorrelated in the first two moments and $\mathbb{E}[z_t] = 0$, $\mathbb{E}[z_t^2] = 1$, if the model is correctly specified (see for example Jung et al. 2006, p. 2360, Liesenfeld et al. 2008, p. 190).

A test for serial correlation which is regularly used for ARIMA modelling, but asymptotically also valid for count data models, when residuals fulfill above assumptions, is the so-called *Ljung-Box portmanteau test* (Cameron and Trivedi 2013, p. 270). It is based on the autocorrelation function, which was already defined above for observed values and their lags, and is now set up using the Pearson residuals in (15):

$$\hat{\rho}_q = \frac{\sum\limits_{t=q+1}^{T} z_t \, z_{t-q}}{\sum\limits_{t=1}^{T} z_t^2} \quad . \tag{16}$$

The autocorrelation coefficient is now marked as an estimated parameter, because the Pearson residuals are obtained after fitting an appropriate model (ibid.).

The denominator simplifies compared to equation (7), because the residuals are assumed to be homoskedastic for a correctly specified model. The test evaluates whether either the autocorrelation coefficients up to a predetermined lag order k are jointly equal to zero, $H_0 : \rho_q = 0,\ q = 1, \dots, k$, or at least one coefficient is statistically different from zero. In order to do so, the Ljung-Box test statistic is defined as

$$\mathrm{T_{LB}} = T\,(T+2) \sum_{q=1}^{k} \frac{1}{T-q}\,\hat{\rho}_q^2 \tag{17}$$

which, under H_0, is asymptotically $\chi^2(k)$ (Cameron and Trivedi 2013, p. 270). A level of significance $\alpha = 0.05$ is considered as the benchmark when testing for serial correlation, as for the t-test mentioned earlier.

In addition to the model evaluation based on single observations using the Pearson residual, overall performance of the model is of interest, especially when it comes to compare competing models. A commonly used tool for this purpose is the *Pearson statistic* which is based on the Pearson residuals and serves as a measure for the goodness-of-fit of the model (Cameron and Trivedi 2013, pp. 188). The Pearson statistic is essentially a weighted sum of residuals, given by

$$P = \sum_{t=1}^{T} \frac{(y_t - \hat{\mu}_t)^2}{\sqrt{\hat{\omega}_t}} \quad , \tag{18}$$

where all elements of the equation are defined as for the Pearson residual above (ibid.). This statistic is a helpful tool for both the Poisson and the Negbin II model to assess their specification and to detect possible overdispersion. For a well specified model and under the assumption of a correctly specified mean, it holds that $P = n - k$, whereas $P > n - k$ and $P < n - k$ indicate over- and underdispersion, respectively[2] (ibid.). Note, however, that given the assumption of a correctly specified mean, $P \neq n - k$ can either indicate a misspecification of the distributional assumptions or the mean itself (Winkelmann 2008, p. 119).

Two further criteria for model evaluation and comparison of non-linear models should not be neglected, namely the *Akaike information criterion* (AIC) and the *Bayesian information criterion* (BIC). Having in mind, that a forecasting model should include as much information as possible from the original time series and be parsimoniously parametrized at the same time, these information criteria provide a quantification of this trade-off (González-Rivera 2013, pp. 212). The

[2] This stems from the properties of the Pearson residual: For correct specification, $\mathbb{E}[(y_t - \mu_t)^2/\omega_t] = 1$, which implies $\mathbb{E}[\sum_{t=1}^{T}(y_t - \mu_t)^2/\omega_t] = n$. The degrees of freedom adjustment $(n - k)$ is made, because in practice μ_t has to be estimated (Winkelmann 2008, p. 119).

intuition behind the criteria is, that on the one hand, an increase in the number of parameters k in the model improves its goodness-of-fit, since more information can be captured, but violates the principle of parsimony on the other hand. Translating this intuition into equations yields

$$AIC = \frac{2k}{T} + \ln \frac{\sum_{t=1}^{T} z_t^2}{T} \tag{19}$$

$$BIC = \frac{k}{T} \ln T + \ln \frac{\sum_{t=1}^{T} z_t^2}{T} \quad , \tag{20}$$

where z_t is the Pearson residual defined in (15) and T is the number of observations (González-Rivera 2013, pp. 212). The first term of both criteria is a penalty term which penalises each parameter that is added to the model, the second term includes the residual sum of squares which is assumed to decrease as parameters are added to the model, so that both terms counterbalance each other. The decision rule is to choose the model with the smallest AIC or BIC among the competing models. Both criteria have different optimality properties and there is no consensus of preferring one over the other, so both are taken into account in this paper (ibid.).

3.3 Model-based Forecasting

After this review concerning the present and basic statistical methods dealing with it, this section treats the future and the therefor appropriate statistical method, which is model-based forecasting. Following the definitions given above, the aim of model-based forecasting is to construct a forecast, $f_{t,h}$, for some future realisation y_{t+h} of the stochastic process $\{Y_t\}_{t=1}^{T}$ by exploiting all information available up to time t, $I_t = \{y_1, y_2, \dots, y_t, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t\}$, which is called a *multivariate information set* (González-Rivera 2013, p. 73). This forecast is therefore defined as a function $g(\cdot)$ of the information set,

$$f_{t,h} = g(I_t), \tag{21}$$

where $g(\cdot)$ is a function of conditional moments (ibid.). Since this paper tries to forecast the number of emergency arrivals, which is modelled in terms of a regression framework and hence in terms of the conditional mean, $g(\cdot)$ is a function of the conditional mean and so is the forecast. Then, (21) can be rewritten as

$$f_{t,h} = g(I_t) = \mathbb{E}[Y_{t+h}|I_t] = \mathbb{E}[Y_{t+h}|y_1, y_2, \dots, y_t, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t] \tag{22}$$

and it turns out that this is an expression well calculable using the specification in (10):

$$\begin{aligned} \mathbb{E}[y_t|\mathbf{x}_t, \mathbf{y}_{t-q}] &= exp(\mathbf{x}_t'\boldsymbol{\beta} + \mathbf{y}_{t-q}'\,\boldsymbol{\rho}_q) \\ f_{t,h} = \mathbb{E}[Y_{t+h}|\mathbf{x}_t, \mathbf{y}_{t+h-q}] &= \mu_{t+h|h} = exp(\mathbf{x}_t'\boldsymbol{\beta} + \mathbf{y}_{t+h-q}'\,\boldsymbol{\rho}_q). \end{aligned} \tag{23}$$

Thus, just as the conditional mean is a log-linear function of regressors and lagged dependent variables, the forecast is a log-linear function of past observations which is necessary for applying the standard forecasting theory presented in Hamilton (1994, Ch. 4). Additionally, (23) implies, that Poisson and Negbin II models yield the same forecast, since their conditional mean is specified in the same way. For obvious reasons, the *forecast error* is defined using (23) as the difference between the actual observation at time $t + h$ and the forecast (González-Rivera 2013, p. 11),

$$e_{t,h} = y_{t+h} - f_{t,h}. \tag{24}$$

Clearly, even given the information in (24), no statement about the usefulness of the theoretical forecast in (23) is possible so far. To be able to assess and compare different forecasts, some additional choices have to be made *a priori*, that is before constructing the forecast (González-Rivera 2013, p. 79).

First, the nature of the information set has to be clarified, whether it is univariate or multivariate, quantitative or qualitative, because the forecast is always a function of this information set as pointed out in (21) (González-Rivera 2013, pp. 80). Second, the *forecast horizon* must be predetermined, that is, whether a one-step ahead forecast, $f_{t,1}$, or a multi-step ahead forecast, $f_{t,h}, h > 1$, will be constructed. Another differentiation would be according to long-, medium-, or short-term forecasts, however, this strongly depends an the frequency of the data[3] (González-Rivera 2013, pp. 84). Related to the issue of the forecast horizon is the choice of a *forecasting environment*. Since it is impossible to observe the future $t + h$ at time t, where the forecast is constructed and its performance has to be assessed, the sample of observations $\{y_1, y_2, \ldots, y_T\}$ is divided into an *estimation sample*, $\{y_1, y_2, \ldots, y_t\}$, which is used to estimate the econometric model and a *prediction sample*, $\{y_{t+1}, y_{t+2}, \ldots, y_T\}$, which is used to assess the forecast performance. This makes it possible to calculate the forecast error as given in (24) which is known as *out-of-sample* assessment (González-Rivera 2013, p. 86). The forecasting literature distinguishes three different forecasting environments, namely the *recursive*, the *rolling* and the *fixed* (for a detailed discussion see González-Rivera 2013, pp. 86). Using the recursive forecasting environment, the model is estimated inside the estimation sample up to time t and the forecast, for example $f_{t,1}$, is made for $t + 1$ afterwards. Subsequently, the estimation sample is expanded until $t + 1$, the model is estimated again and the one-step ahead forecast is constructed for $t + 2$. This procedure continues until no observations are left in the prediction sample. In general, the recursive forecasting environment is advantageous, because it uses the maximum of information to estimate the model. However, its validity relies on the stability of the model over time (ibid.). Since this is a reasonable assumption for a model of emergency arrivals, this forecasting environment is applied in the practical part of this study. Note, that producing forecasts using the recursive scheme also leads to a series of out-of-sample forecast errors, in the example of one-step ahead forecasts $\{e_{t,1}, e_{t+1,1}, \ldots, e_{t+j,1}, \ldots, e_{T-1,1}\}$ or, for the general case, $\{e_{t,h}, e_{t+1,h}, \ldots, e_{t+j,h}, \ldots, e_{T-h,h}\}$. These series can be assessed by computing the *sample average loss*, which can be expressed by different measures (González-Rivera 2013, p. 239 presents the most common measures). For reasons which will be pointed out in an instant, the sample average loss in this paper is assessed

[3] Think, for example, of the difference between quarterly reported data on the gross domestic product and stock market data which is nearly continuously reported every fraction of a second.

by using the *Mean Squared Error* (MSE), defined as

$$\bar{L} = \frac{\sum_{j=0}^{T-h-t} e_{t+j,h}^2}{T-h-t+1} \equiv \text{MSE}, \tag{25}$$

and the root mean squared error, RMSE $= \sqrt{\text{MSE}}$. The model producing forecasts with the smallest MSE will be preferred to possible alternatives. Another possibility to evaluate the performance of forecasts is the concept of a *loss function* (González-Rivera 2013, pp. 89). This is the last *a priori* choice to be made by the forecaster. A loss function $L(e_{t,h})$ is defined as "the evaluation of costs associated with the forecast errors" (González-Rivera 2013, p. 90), or, in everyday language, "how concerned we are if our forecast is off by a particular amount" (Hamilton 1994, p. 72). Regardless of the functional form of $L(e_{t,h})$, every loss function must satisfy three conditions which are also presented in González-Rivera (2013, pp. 90):

- $e_{t,h} = 0 \quad \rightarrow \quad L(e_{t,h}) = 0$
- $e_{t,h} \neq 0 \quad \rightarrow \quad L(e_{t,h}) \geq 0,\ min\ L(e_{t,h}) = 0$
- if $e_{t,h}^{(1)} > e_{t,h}^{(2)} > 0 \quad \rightarrow \quad L(e_{t,h}^{(1)}) > L(e_{t,h}^{(2)})$

  if $e_{t,h}^{(1)} < e_{t,h}^{(2)} < 0 \quad \rightarrow \quad L(e_{t,h}^{(1)}) > L(e_{t,h}^{(2)})$.

Expressing these conditions in words, the loss function is a non-negative function, which takes on its minimum value zero when the forecast equals the observed value and is monotonically increasing for positive and monotonically decreasing for negative errors, respectively. It is up to each forecaster to specify her loss function according to her own preferences and various examples for imaginable loss functions exist (again, a very comprehensive treatment is provided in González-Rivera 2013, pp. 91). Albeit being the most unrealistic specification of a loss function in the majority of applications, the quadratic loss function $L(e_{t,h}) = a\, e_{t,h}^2,\ a > 0$ is the most widely used, because it can be handled very conveniently from a mathematical point of view (Hamilton 1994, pp. 72). Verbalising this definition, it can be said that the sign of the forecast error is irrelevant, since it is squared and therefore non-negative, which is necessary to fulfill the first and the second condition above. Additionally, and this is the crucial assumption made by using a *symmetric* loss function, forecast errors of the same magnitude are weighted equally, regardless whether they are positive or negative (González-Rivera 2013, pp. 91).

3.3.1 Optimal Forecasting

With the *a priori* choices of the information set, the forecast horizon and environment, and especially the loss function being made, it is possible to derive the *optimal forecast*, denoted by $f^*_{t,h}$. González-Rivera (2013, pp. 93) illustrates an intuitive and basic derivation, a more general, yet technically more involving, derivation can be found in Hamilton (1994, p. 73). The former is explicitly presented below, since some remarks have to be made on the case of a discrete probability mass function like the Poisson or the negative binomial.

From the definition of the forecast error in (24) and under the assumption of symmetric loss, the loss function can be written as

$$L(e_{t,h}) = L(y_{t+h} - f_{t,h}) = a\, e^2_{t,h} \quad , \tag{26}$$

where y_{t+h} is a future realisation of some random variable Y_{t+h}. Now, whereas González-Rivera (2013, p. 94) assumes this random variable to have a normal conditional density, this is not the case in this study, where the emphasis is on models for count data and the conditional density $f(y_{t+h}|\mathbf{x}_t, \mathbf{y}_{t-q}; \boldsymbol{\theta})$ has to be expressed in terms of expressions (8) and (12). In any case, Y_{t+h} being a random variable with unknown realisation y_{t+h} leads to $L(\cdot)$, which in turn is a random variable as well. That makes it possible to examine its statistical moments, where the expected loss is of particular interest (ibid.). For the case of a discrete random variable Y_{t+h}, the expected loss can be expressed as

$$\mathbb{E}[L(y_{t+h} - f_{t,h})] = \int_0^\infty L(y_{t+h} - f_{t,h})\; f(y_{t+h}|\mathbf{x}_t, \mathbf{y}_{t-q}; \theta) \tag{27}$$

and it turns out, that this is nearly the same expression as presented in González-Rivera (2013, p. 94) for the continuous case. This stems from the fact, that the loss function is a function of the difference between the future realisation y_{t+h}, which is a count for a discrete probability distribution, and the forecast $f_{t,h}$, which can be expressed as a conditional mean specified within a regression framework as shown in (23). Thus, the same domain as for the conditional mean applies to the forecast and it is $f_{t,h} \in \mathbb{R}^+$. This implies, that the outcome of the loss function is continuous as well and standard rules of integration and differentiation can be applied. The optimal forecast is now defined as the forecast which “minimises the expected loss” (González-Rivera 2013, p. 94),

$$\min_{f_{t,h}} \mathbb{E}[L(y_{t+h} - f_{t,h})]. \tag{28}$$

The minimisation problem can be solved by plugging in the quadratic loss function from above and using simple rules of differentiation which can also be found in González-Rivera (2013, p. 94) and is not repeated here. It turns out, that the optimal forecast equals the conditional mean of y_{t+h},

$$f^*_{t,h} = \mathbb{E}[Y_{t+h}|\mathbf{x}_t, \mathbf{y}_{t+h-q}] \equiv \mu_{t+h|t}. \tag{29}$$

It has to be stressed, that this expression heavily depends on the choice of the loss function and is therefore only valid for the quadratic loss function. Furthermore, it becomes obvious now, why the (R)MSE is used in this paper to assess the sample average loss, because it is closely related to the definition of the quadratic loss function: While the optimal forecast minimises the expected loss as shown in González-Rivera (2013, p. 94), this is also true for the (R)MSE as shown in Hamilton (1994, p. 73).

3.3.2 Constructing and Evaluating Density Forecasts

Until now, only so-called *point forecasts* were discussed, where the forecast returns a single value which is considered as a prediction for a future realisation of a random variable (González-Rivera 2013, p. 12). Point forecasts are in general valuable in giving a rough idea of how a time series might develop in the future, yet they do not provide any measure of probability on how likely it is that the forecast in fact equals the observed future value. There is more information which can be provided and which is particularly valuable for practitioners. One way to do so is the construction of *density forecasts*, exploiting the fact that the time series which is forecast is assumed to be governed by a stochastic process, that is, a collection of random variables, where each of them possesses its specific probability mass function. Hence, the density forecast is defined as the probability mass function of Y_{t+h} conditional on the information set I_t, $f_{t,h}(Y_{t+h}|I_t)$ (ibid.). This definition indicates, that, albeit leading to the same point forecasts, the Poisson and the Negbin II model yield different density forecasts.

For the Poisson model, the density forecast is easily obtained: As pointed out in the last section, the optimal forecast under quadratic loss is the conditional mean $\mu_{t+h|t}$, which is also the parameter governing the conditional probability mass function of the Poisson model. The Poisson density forecast can then be written as

$$f_{t,h}(Y_{t+h}|\mathbf{x}_t, \mathbf{y}_{t+h-q}; \boldsymbol{\theta}) = \frac{e^{-\mu_{t+h|t}} \mu_{t+h|t}^{y_t}}{y_t!} \quad , \tag{30}$$

where now $\boldsymbol{\theta} = \{\mu_{t+h|t}\}$. For the Negbin II model, the density forecast is obtained in a similar way, using the probability mass function in (12) and plugging in $\mu_{t+h|t}$, which yields

$$\begin{aligned} &f_{t,h}(Y_{t+h}|\mathbf{x}_t, \mathbf{y}_{t+h-q}; \boldsymbol{\theta}) \\ &\quad = \frac{\Gamma(\alpha + y_{t+h})}{\Gamma(\alpha)\Gamma(y_{t+h} + 1)} \left(\frac{\alpha}{\alpha + \mu_{t+h|t}}\right)^{\alpha} \left(\frac{\mu_{t+h|t}}{\alpha + \mu_{t+h|t}}\right)^{y_{t+h}}. \end{aligned} \tag{31}$$

So, by constructing the density forecast, an entire probability distribution of possible future outcomes is estimated and the entire forecast uncertainty is depicted (Tay and Wallis 2000, p. 236). Again, just as point forecasts, density forecasts have to be assessed in order to choose the model with the best predictive ability. However, while there exists a rich literature on evaluating point forecasts, this is not the case for the evaluation of density forecasts (Diebold et al. 1998, p. 863). In this study, a very straightforward approach to reach the objective is used. The strategy is to borrow from the econometric field of kernel density estimation, where different measures for the global performance of estimated densities have been developed in the literature. For a density forecast $f_{t,h}(Y_{t+h}|I_t)$ and a true density at time $t+h$, $f_{t+h}(y_{t+h}|\mathbf{x}_t, \mathbf{y}_{t+h-q})$, a commonly used measure of global performance is the *Mean Integrated Squared Error* (MISE), defined as

$$MISE(f_{t,h}) = \int_{-\infty}^{\infty} \{f_{t,h}(z) - f_{t+h}(z)\}^2 \pi(z) dz, \tag{32}$$

where z are the arguments of the probability density functions which are defined for the real numbers and $\pi(z)$ is a weight function which is set equal to one from now on (Tenreiro 1998, p. 134). This measure can be seen as the counterpart of the MSE for estimated and forecasted densities. However, for the application to the density forecasts of discrete probability distributions, the MISE is inappropriate, since it relies on continuous integration over the real numbers.

Wegman (1972, p. 228) presents a discrete version of the MISE, the *Average Square Error* (ASE), given by

$$ASE = \frac{1}{N} \sum_{n=1}^{N} \{f_{t,h}(z) - f_{t+h}(z)\}^2, \tag{33}$$

where integration is replaced by summation over the positive integers and z is the argument of the density functions, which can now be replaced by discrete probability mass functions (ibid.). This easily-computable measure returns the average squared deviation between the density forecast and the assumed true density, so $\sqrt{ASE}$ returns the average absolute deviation. When comparing the predictive ability of different forecasting models for emergency arrivals below, the ASE for each density forecast $f_{t,h}(Y_{t+h}|I_t)$ in the prediction sample is computed and the overall mean of the ASEs is reported as a measure $\overline{ASE}$ to evaluate the predictive ability of a forecasting model. Naturally, the model with the smallest $\overline{ASE}$ will be preferred over alternative models to forecast the emergency arrivals of a certain priority.

4 Practice: Modelling the Arrival Counts

Having set the stage by discussing the theoretical framework for the present study, this section makes the attempt to apply this framework to a dataset of emergency arrivals in order to develop a model which helps to predict future arrivals. As outlined in the introduction, the procedure is divided into three stages and follows the approach for ARIMA modelling and forecasting presented in González-Rivera (2013, p. 202):

- The data: source, definition, descriptive statistics and autocorrelations.
- The model: estimation, evaluation and selection.
- The forecast: selection of a loss function and construction of the forecast.

All statistical analyses were processed using the software package Stata, release 13.

4.1 The Data

The data under consideration is a sample provided by the emergency department of a hospital in Brandenburg, Germany, and reaches from 1st January 2013 until 30th September 2014 (638 days) on a minute-by-minute basis which gives a total of 61,193 arrivals. Additionally, each emergency patient is assigned to one out of four groups according to the time which is allowed to elapse before medical treatment for the patient starts. These groups can be considered as an approximation of a patient's acuity level and are defined as

- priority 1 (P1): immediately,
- priority 2 (P2): $<$ 30 minutes,
- priority 3 (P3): $<$ 60 minutes and
- priority 4 (P4): $<$ 120 minutes.

In order to develop a forecasting model which produces predictions on a daily basis, emergency arrivals belonging to the same date were aggregated to generate a time series with $N = 638$ observations, equal to the number of days T in the sample. This procedure was performed for the time series of each single priority, so in total four distinct time series were generated. To enable an out-of-sample evaluation of forecasts, the sample was divided into the estimation and the prediction sample. Following the recommendations in González-Rivera (2013, p. 231), the initial division was made at 31st July 2014, so that 577 observations remained in the estimation sample, whereas 61 observations - or 9.56% - of the entire sample were left in the prediction sample. Table 1 illustrates

	overall	P1	P2	P3	P4
Mean	95.54	5.08	53.68	33.94	2.84
Median	95	5	54	33	2
Minimum	61	0	31	10	0
Maximum	147	13	104	65	16
Variance	197.92	6.07	94.25	76.67	5.67
Skewness	0.46	0.46	0.50	0.34	1.5
N	577	577	577	577	577

Table 2: Descriptive statistics for the time series of emergency arrivals

descriptive statistics for the estimation sample of each time series, revealing that P2 and P3 account for the majority of emergency arrivals, while the proportion falling into P1 and P4 is relatively small. Remarkably, all series are positively skewed, which points towards the application of models based on the probability distributions presented earlier. Also striking is the fact, that for all time series the unconditional variance exceeds the unconditional mean, which can serve as a first indicator for possible overdispersion in the data.

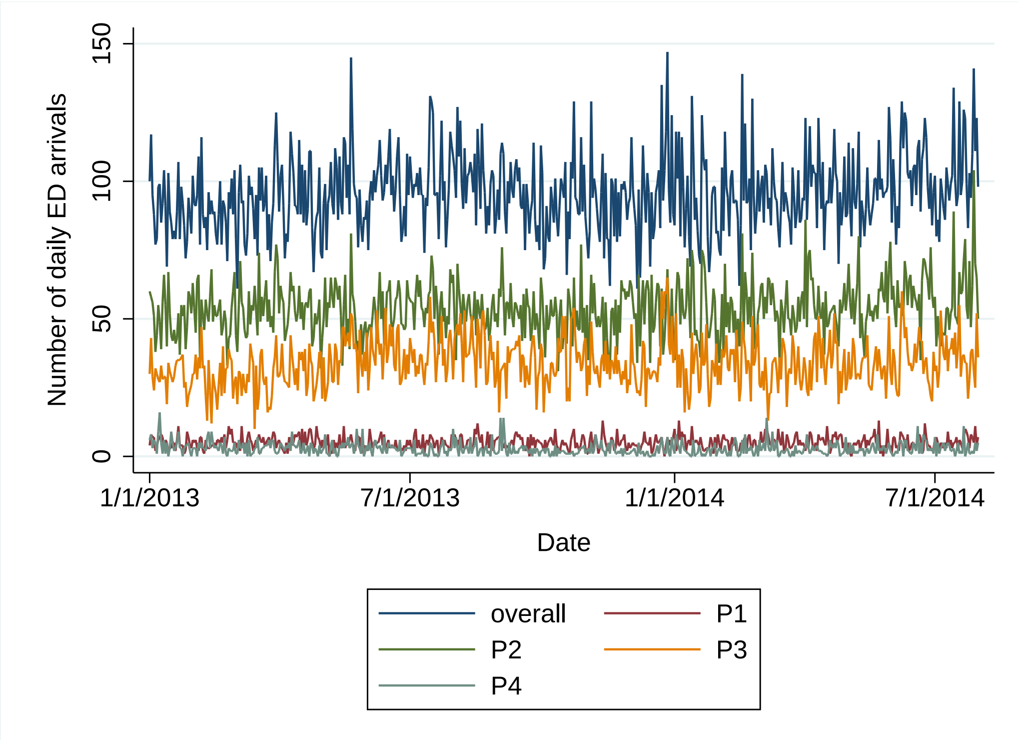

Figure 2: Time series of emergency arrivals

Graphical inspection of the time series depicted in Figure 2 shows, that they are not trending over time, but reveal some sort of cyclical behaviour with a higher number of arrivals in summer, visible as smooth waves in the time series. The existence of a seasonal pattern coincides with other results presented in the literature and discussed in chapter two, however, these results point towards a peak during the winter months and less emergency arrivals in summer. This opposite finding might at least partly stem from the fact, that there are several swimming lakes located in striking distance of the hospital, where swimming accidents can occur during the summer months. A weekly cycle or day of the week effect, another result presented by already existing investigations, cannot be found thoroughly in the sample. Although the overall time series and also P2 reveal a weekly cycle to some extent, visible as zigzag movement of the time series with many spikes, there is only weak evidence for this result from the graphs of P1, P3 and P4.

Following the approach of McCarthy et al. (2008), the raw autocorrelation inherent in the time series is calculated using the formula for the ACF from above, up to the emergency arrivals 28 days ago. The results are depicted in the form of autocorrelograms in Figure 3, where the grey areas correspond to the 95 %-confidence interval of the estimated autocorrelation coefficient $\hat{\rho}_q$, the magnitude of the coefficient is shown on the ordinate and the lag-order q is shown on the abscissa.

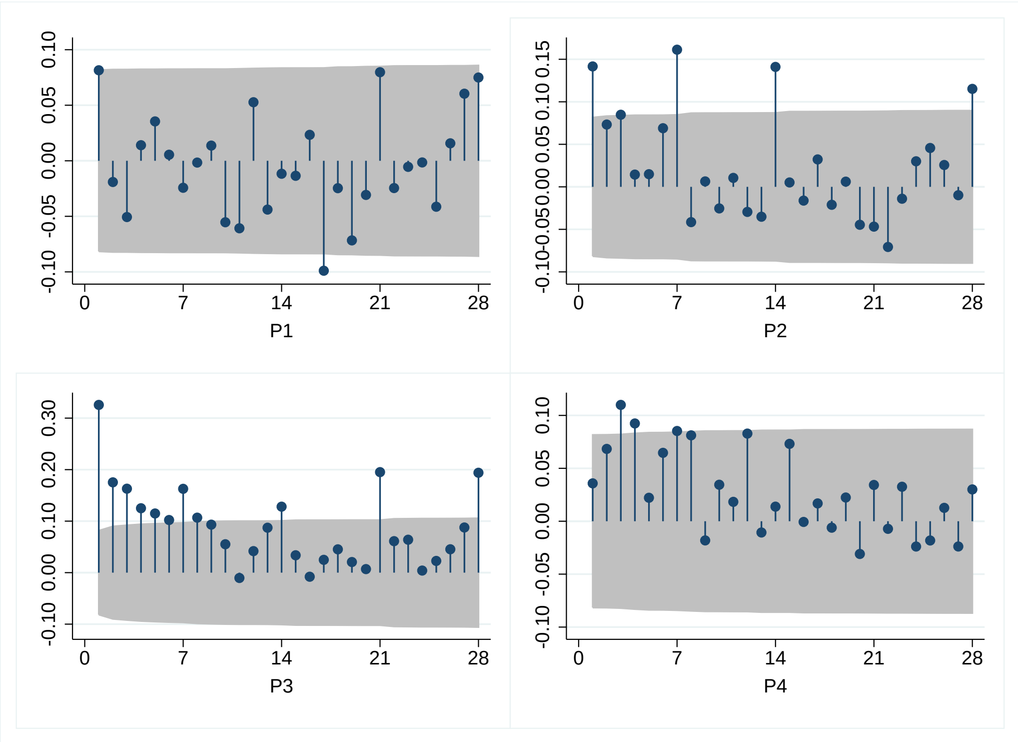

Figure 3: Raw autocorrelation functions

It becomes evident, that there exists a dependence structure in the time series and again, this is especially the case for P2 and P3, whereas the autocorrelation coefficients for P1 and P4 are both less pronounced and rather borderline significant. The time series P2 and P3 exhibit stronger autocorrelation, in particular for the lags $q = 1, 7, 14, 21, 28$ which are - except for P2, $q = 21$ - entirely different from zero. This in fact points towards a weekly cycle in the emergency arrivals, where there is some correlation between arrivals of a specific day of the week and the arrivals seven days apart, which obviously is the same day of the week. Another observation from the autocorrelograms are the exclusively positive autocorrelations between present emergency arrivals and those several days apart.

To pick up the idea brought up above, the positively skewed distributions of emergency arrival counts are compared to the Poisson and the negative binomial distribution, respectively, to gain a first insight of whether these probability distributions can appropriately fit the data. For that purpose, the distribution of emergency arrivals for each priority is depicted as a blue line in Figure 4, along with the corresponding Poisson (green line) and negative binomial distribution (red line). The abscissae show the emergency arrival counts and on the ordinates the respective proportion is shown, that is, the fraction of days on which this exact number of emergency arrivals occurred. Note, that all distributions, while being of discrete nature, are drawn as if they were continuous for reasons of comparability.

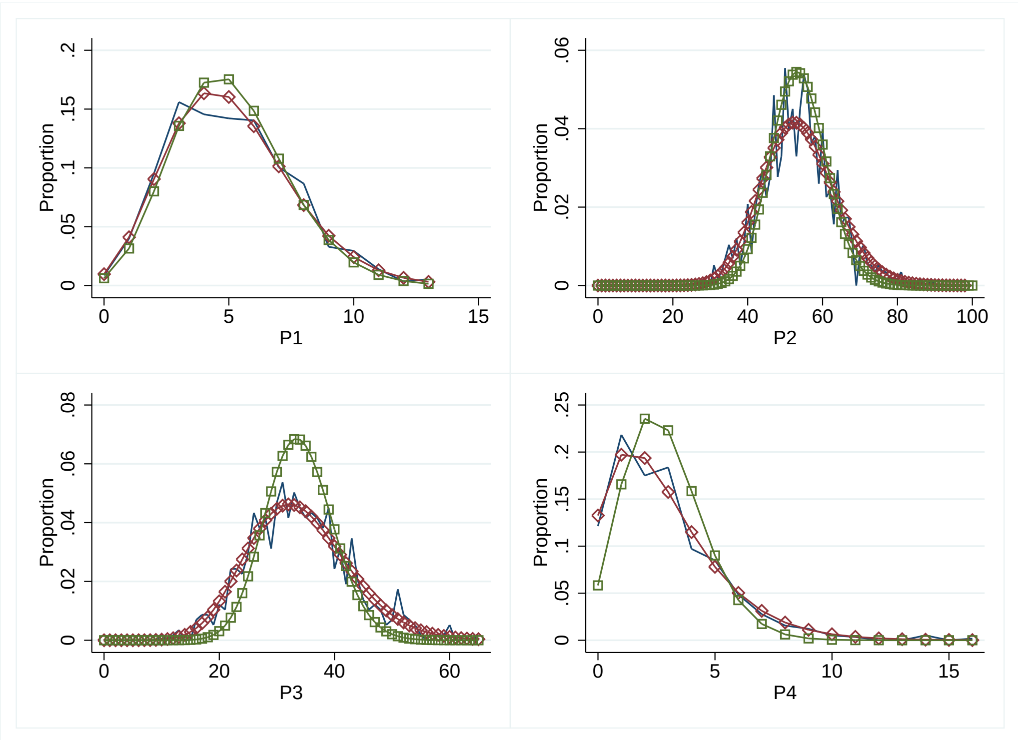

Figure 4: Distributions of emergency arrivals compared to Poisson/Negbin distribution

The graphs show, that the Poisson and negative binomial distribution indeed capture the distributions of emergency arrivals well for all priorities, which makes it reasonable to use them as a basis for the forecasting model and to assume the true data generating process to follow either a Poisson or a negative binomial distribution. Another hint for the modelling procedure, which has to be taken into account, is the observation, that particularly for P3 and P4, the negative

Variable	Description
$TEMP$	daily average air temperature (°C)
$TEMPMAX$	daily maximum air temperature (°C)
$TEMPMIN$	daily minimum air temperature (°C)
$RAIN$	rainfall (mm)
$SNOW$	snowfall (cm)
HUM	relative humidity (%)
$VAPRESS$	vapour pressure (hPa)
$D_{TEMPMAXi}$	variable = 1 for $TEMPMAX \geq i$, $i = 0$, 10, 15, 20, 30 °C, 0 otherwise
D_{RAINi}	variable = 1 for $RAIN > i,\ i = 1,\ 15$ mm, 0 otherwise
D_{SNOWi}	variable = 1 for $SNOW > i,\ i = 1,\ 5,\ 10$ cm, 0 otherwise
$D_{VAPRESSi}$	variable = 1 for $VAPRESS \geq i$, $i = 10,\ 15,\ 20$ hPa, 0 otherwise

Table 3: Climatic variables used for modelling

binomial distribution is closer to the observed data than the Poisson distribution.

To sum up the insights obtained from the raw time series, their graphical representations and descriptive statistics, few commonalities and a high degree of heterogeneity between priorities could be found. In general, all time series are not trending over time, the unconditional variance exceeds the unconditional mean and the distributions of emergency arrival counts are positively skewed. However, the priorities differ remarkably with regard to their sheer magnitude, their cyclicality and their inherent dependence structure. These results make it reasonable to develop four separate forecasting models, one for each priority, as suggested by Sun et al. (2009), and build them on grounds of probability models for count data.

One further step, which has been made to enable econometric modelling using a regression framework and to provide as much information as possible to verify results which already exist, was to merge the dataset of emergency arrival counts with climate data obtained from a meteorological station eight kilometers away from the hospital (DWD 2015). This dataset with daily observations contains various climatic variables such as the daily average air temperature, the daily maximum and minimum air temperature, the relative humidity, the vapour pressure and the amount of precipitation reported for rain and snow separately. To simplify the modelling procedure and to obtain variables which indicate the existence or inexistence of a certain event, various binary or dummy variables were defined. They can be found in Table 2 along with the other climatic variables used for econometric modelling. The same step of transforming information into

Variable	Description
$EA_t^{(i)}$	emergency arrivals for priority $i = 1, 2, 3, 4$ at time $t = 1, \ldots, 577$
D_{iDAY}	variable = 1 for i = MON, TUE, WED, THU, FRI, SAT, SUN, 0 otherwise
$D_{(af)PH}$	variable = 1 for day being public holiday (day after public holiday), 0 otherwise
D_{VAC}	variable = 1 for day being part of school holidays, 0 otherwise
D_{MONTH}	variable = 1 for MONTH = JAN, FEB, MAR, APR, MAY, JUN, JUL, AUG, SEP, OCT, NOV, DEC, 0 otherwise
D_{SUMMER}	variable = 1 for MONTH = JUN, JUL, AUG, 0 otherwise
D_{WINTER}	variable = 1 for MONTH = OCT, NOV, DEC, 0 otherwise

Table 4: Further variables used for modelling

variables which can be analysed within a regression framework was also made with regard to the 577 different dates in the estimation sample. They were characterised by the month in which they fall, their specific day of the week and the nature of the day, that is, whether it is a public holiday, the day before or after a public holiday or whether it is part of school holidays. Additionally, the months June, July and August were used to define a variable which indicates whether it is summer or not, the same procedure was carried out for October, November and December being defined as winter. The entire process of defining variables was performed by creating dummy variables, indicating whether a certain event is true or not for a specific date in the time series by taking on the values one or zero (for more information on the use of binary variables see Wooldridge 2013, Ch. 7). The list of variables other than the ones already presented in Table 2 can be found in Table 3 along with their respective definitions.

4.2 The Model

In order to develop a forecasting model for the emergency arrival counts of the different priority groups, the conditional means of the Poisson and the negative binomial distribution are specified in terms of a regression model as shown in the theoretical part of this paper. That allows to relate the dependent variable $EA_t^{(i)}$, todays emergency arrivals, to its previous values $EA_{t-q}^{(i)}$ and to control for various effects, for instance seasonality, the day of the week or temperature. The modelling strategy follows a general-to-specific approach as suggested by Dougherty (2011, pp. 457), starting with the most general model, where both exogenous regressors and lagged values of the dependent variable enter the regression equation to incorporate as much information as possible and to capture all dependence in the time series within the specified model, so that the Pearson residuals have mean zero, variance one and no serial correlation is left in z_t and z_t^2. Afterwards, non-significant parameters are dropped from the general model to simplify it if possible. Before starting the estimation step, some additional remarks seem to be appropriate: All models are estimated using either Poisson QML estimation or Negbin II ML estimation and including a constant, so that dummy variables are given the interpretation of a difference compared to a reference or base group which is not part of the model (Wooldridge 2013, p. 227). Estimated coefficients marked as (*) in the following results are statistically significant on the 5%-level, for coefficients marked as (**) the same is true even on the 1%-level. For the Ljung-Box portmanteau statistic, T_{LB}, the corresponding p-value is reported in brackets. For reasons of brevity, the final equation for the forecasting models is only presented for P1 as an example, because the equations for P2-4 follow exactly the same structure but contain too many parameters in most of the cases. Note, that for the same reason only the *best* models among those which were estimated are presented below. The absence of explosiveness in the models due to $\boldsymbol{\rho}_q > 1$ was tested by plotting the fitted values from the regression and could be confirmed for all of the models.

4.2.1 Priority 1

As evident from the graphical representations and descriptive statistics of the time series, the number of patients classified into P1, $EA_t^{(1)}$, is rather small, compared to other priority levels and as a fraction of the overall time series of emergency arrivals. Additionally, Figure 2 does not indicate strong cyclical behaviour of the time series, neither on a weekly, nor on a yearly basis and according to Figure 3, present emergency arrivals of P1 are only weakly correlated with previous arrivals one and 17 days ago. These observations can serve as a first indication for the difficulty to predict the outcomes of this series. As suggested by aforementioned observations, all climatic and monthly predictors from Tables 2 and 3 in fact turn out to have no statistically significant effect on $EA_t^{(1)}$, visible in p-values far above .05. The same holds for the effect of a day being a public holiday or part of school holidays and the effects of summer and winter. The only deterministic predictor which could be identified is D_{MONDAY} with a p-value of $.003 < .05$. As illustrated in Table 4, $EA_t^{(1)}$ can also be predicted by its own lags $EA_{t-1}^{(1)}$ and $EA_{t-17}^{(1)}$, which is the stochastic part of the model, since the lags are realisations of the random variable $EA_t^{(1)}$.

The postestimation procedures reveal, that there is no correlation left in the regression residuals, visible in the p-value corresponding to the Ljung-Box portmanteau test of .50, which is far from rejecting the null hypothesis of no serial correlation. Furthermore, the residuals show the desirable property of a mean near to zero and a variance close to one. This provides some evidence that the model is well specified. However, the Pearson statistic is slightly larger than

Poisson model		Negbin II model	
$EA_t^{(1)}$	Coefficient	$EA_t^{(1)}$	Coefficient
D_{MONDAY}	.1530**	D_{MONDAY}	.1526**
$EA_{t-1}^{(1)}$	.0175*	$EA_{t-1}^{(1)}$	.0176*
$EA_{t-17}^{(1)}$	-.0204**	$EA_{t-17}^{(1)}$	-.0207*
constant	1.6147	constant	1.6153
$\hat{\alpha}$	-	$\hat{\alpha}$	.0343**
z_t		z_t	
mean	-.0001	mean	.0000
std. err.	1.0877	std. err.	.4440
T_{LB}	39.1657 (0.50)	T_{LB}	40.1397 (0.46)
P/df	1.1895	P/df	1.1895
AIC[5]	2572.422	AIC	2565.019
BIC	2589.734	BIC	2582.331

Notes: (*) indicates statistical significance on the 5%-, (**) on the 1%-level.

Table 5: Regression results for P1

$(n - k)$, $661.37 > 577 - 17 - 4$, which points towards some degree of overdispersion[4]. Even though this does not affect the estimated regression coefficients, a Negbin II regression model was estimated as well, using QML estimation to compare the diagnostic results to those obtained from the Poisson model. The regression residuals for the Negbin II model do not show any correlation and the information criteria of the models are hardly discriminable. The only visible difference is the size of the residual variance for the Negbin II model, which is $.4440^2 = .1971$ and therefore closer to zero than to one, where it should be for a well specified model. After the estimation and evaluation step, there are two possible model candidates for P1, namely the Poisson or the Negbin II specification, containing the same parameters - because their conditional means are specified the same in theory - and resulting in nearly the same diagnostic results. For both models, the estimated conditional mean with parameter estimates rounded up to two decimal places is specified as

$$\mathbb{E}[EA_t^{(1)}|D_{MONDAY}, EA_{t-1}^{(1)}, EA_{t-17}^{(1)}] = \mu_t^{P1}$$

$$= exp(1.61 + 0.15 \cdot D_{MONDAY} + 0.02 \cdot EA_{t-1}^{(1)} - 0.02 \cdot EA_{t-17}^{(1)}). \quad (34)$$

[4] Note, that the time series contains 17 dates with $EA_t^{(1)} = 0$.

[5] Note, that Stata calculates the information criteria according to $AIC = -2 \ln L(\cdot) + 2k$ and $BIC = -2\ln L(\cdot) + k \ln(N)$.

This equation will be the basis for constructing forecasts in the next step.

4.2.2 Priority 2

The group of emergency arrivals classified into P2, $EA_t^{(2)}$, represents the largest fraction of the entire time series, its median covers 56.84% of the overall time series' median and it exhibits the same wavelike movement with a higher number of arrivals in the summer months. Additionally, the time series exhibits a rather pronounced dependence structure with positive autocorrelation coefficients at lags 1, 7, 14 and 28. This can be an indication towards some weekly seasonality. All in all, the time series for P2 seems to be better predictable by deterministic factors and its own lags, than this is the case for P1. Just as in the case of P1, regression analysis confirms the first impression from analysing descriptive statistics and graphs: Using a Poisson regression model and Monday as reference day, there is strong evidence for all days of the week from Tuesday until Sunday being predictors for $EA_t^{(2)}$ with p-values practically equal to zero and negative coefficients, indicating a peak on Mondays, which is consistent with the literature presented in chapter two. Further predictors are public holidays and the day after a public holiday, where the former has a negative coefficient, the latter a positive one, indicating that less emergency patients arrive on public holidays, but more the day after. Additionally, D_{WINTER} can serve as predictor which states, that for the months October, November and December less emergency patients arrive compared to the rest of the year. This confirms the seasonality hypothesised from the waves in the time series. The autocorrelation hypothesised from the autocorrelogram is meanwhile confirmed by the estimated regression coefficients for $EA_{t-1}^{(2)}$, $EA_{t-7}^{(2)}$ and $EA_{t-21}^{(2)}$, which differ all statistically from zero, at least on the 5%-level and point - just as the day of the week regressors do - towards a weekly cycle in the emergency arrivals of P2. $TEMP$ was also estimated within the model, resulting in a very small and positive coefficient, which differs from zero on the 5%-level and is kept in the model, since it might explain the higher number of arrivals during the summer. With regard to the evaluation step the model performs well with no serial correlation in z_t, which has a mean close to zero and variance close to one, and z_t^2, so the conditional mean seems to be correctly specified. The only issue is the ratio of Pearson statistic to degrees of freedom, $P/df \approx 1.48 > 1$ which might indicate further overdispersion. To analyse the effect of the temperature on $EA_t^{(2)}$ in greater depth, a second Poisson model was estimated, using the same variables as above, but dropping D_{WINTER} and replacing $TEMP$ by $D_{TEMPMAX20}$. The result is an even stronger and positive effect of the temperature variable on the emergency arrivals as illustrated in Table 5.

Poisson model 1		Poisson model 2	
$EA_t^{(2)}$	Coefficient	$EA_t^{(2)}$	Coefficient
$D_{TUESDAY}$	-.1371**	$D_{TUESDAY}$	-.1353**
$D_{WEDNESDAY}$	-.1085**	$D_{WEDNESDAY}$	-.1087**
$D_{THURSDAY}$	-.1589**	$D_{THURSDAY}$	-.1582**
D_{FRIDAY}	-.0995**	D_{FRIDAY}	-.0981**
$D_{SATURDAY}$	-.1491**	$D_{SATURDAY}$	-.1467**
D_{SUNDAY}	-.1597**	D_{SUNDAY}	-.1551**
D_{PH}	-.1112*	D_{PH}	-.1052*
D_{afPH}	.1938**	D_{afPH}	.1951**
D_{WINTER}	-.0450*	D_{WINTER}	-
$TEMP$	.0019*	$TEMP$	-
$D_{TEMPMAX20}$	-	$D_{TEMPMAX20}$	.0548**
$EA_{t-1}^{(2)}$	.0027**	$EA_{t-1}^{(2)}$	.0027**
$EA_{t-7}^{(2)}$	.0016*	$EA_{t-7}^{(2)}$	.0017*
$EA_{t-21}^{(2)}$	-.0027**	$EA_{t-21}^{(2)}$	-.0025**
constant	3.9977	constant	4.0004
z_t		z_t	
mean	.0000	mean	.0000
std. err.	1.2034	std. err.	1.2002
T_{LB}	43.4216 (0.33)	T_{LB}	45.2666 (0.26)
P/df	1.4829	P/df	1.4724
AIC	4060.688	AIC	4055.455
BIC	4121.178	BIC	4111.625

Notes: (*) indicates statistical significance on the 5%-, (**) on the 1%-level.

Table 6: Regression results for P2 (Poisson models)

For reasons of comparison, two Negbin II models were estimated as well, using the same parametrisation as the Poisson models for P2, the results are summarised in Table 6. Comparing the parameter estimates of the Poisson and the Negbin II specification, a difference is hardly visible. However, when looking at the postestimation results, there is evidence that the Negbin II models might fit the data better than the Poisson models do. On the one hand, this stems from the fact, that the estimated dispersion parameter $\hat{\alpha}$ differs statistically from zero. On the other hand, the analysis of regression residuals provides even stronger evidence: For both Negbin II models, they have a mean equal to zero and a variance equal to one and - most remarkably - the ratio P/df is virtually equal

Negbin II model 1		Negbin II model 2	
$EA_t^{(2)}$	Coefficient	$EA_t^{(2)}$	Coefficient
$D_{TUESDAY}$	-.1377**	$D_{TUESDAY}$	-.1359**
$D_{WEDNESDAY}$	-.1085**	$D_{WEDNESDAY}$	-.1085**
$D_{THURSDAY}$	-.1585**	$D_{THURSDAY}$	-.1578**
D_{FRIDAY}	-.0996**	D_{FRIDAY}	-.0979**
$D_{SATURDAY}$	-.1493**	$D_{SATURDAY}$	-.1471**
D_{SUNDAY}	-.1604**	D_{SUNDAY}	-.1555**
D_{PH}	-.1108*	D_{PH}	-.1047*
D_{afPH}	.1929**	D_{afPH}	.1946**
D_{WINTER}	-.0447*	D_{WINTER}	-
$TEMP$	.0019*	$TEMP$	-
$D_{TEMPMAX20}$	-	$D_{TEMPMAX20}$	.0551**
$EA_{t-1}^{(2)}$	.0027**	$EA_{t-1}^{(2)}$	.0027**
$EA_{t-7}^{(2)}$	.0016*	$EA_{t-7}^{(2)}$	.0017*
$EA_{t-21}^{(2)}$	-.0027**	$EA_{t-21}^{(2)}$	-.0025**
constant	3.9997	constant	3.9734
$\hat{\alpha}$	.0082**	$\hat{\alpha}$	.0081**
z_t		z_t	
mean	.0000	mean	.0000
std. err.	1.0014	std. err.	1.0014
$\mathrm{T_{LB}}$	43.5775 (0.32)	$\mathrm{T_{LB}}$	43.5775 (0.32)
P/df	1.0268	P/df	1.0232
AIC	4017.670	AIC	4013.413
BIC	4078.161	BIC	4069.583

Notes: (*) indicates statistical significance on the 5%-, (**) on the 1%-level.

Table 7: Regression results for P2 (Negbin II models)

to one. Additionally, both information criteria are smaller than the ones for the Poisson models, so all in all, the Negbin II specification is superior in modelling the emergency arrivals for P2. Nevertheless, both specifications are used to construct forecasts in the next step, because final evaluation is based on the model's predictive ability. For reasons of brevity, the equations for the forecasting models are not presented here, as they are for P1, because they contain too many parameters.

4.2.3 Priority 3

For P3, being the second largest group of emergency patients after P2, a similar behaviour as for P2 can be hypothesised from graphical inspection of the time series, autocorrelograms and comparison to count data distributions. The time series seems to be autocorrelated to some extent, pointing towards weekly cyclicality, Figure 4 suggests, that the Negbin II specification might fit the data better than the Poisson specification. Performing a regression analysis for the time series of P3 confirms this hypothesis (results in Table 7). Just as for P2, the day of the week is identified as a strong predictor for $EA_t^{(3)}$ with p-values virtually equal to zero and using Monday and Tuesday as the reference now. So again, there is a peak of arrivals at the beginning of the week, followed by a decrease afterwards. D_{PH} is no valid predictor for the time series of P3, instead, the coefficient for D_{afPH} is larger than it is for P2, indicating an increase in emergency arrivals the day after a public holiday. Additionally, there is a sizeable positive effect of $D_{TEMPMAX30}$ on $EA_t^{(3)}$, which might be related to the aforementioned swimming lakes in the area of the hospital. With regard to stochastic elements affecting $EA_t^{(3)}$, the number of arrivals 1, 21 and 28 days ago could be identified, all of them being highly significant and providing evidence of some weekly cyclicality in the emergency arrivals of P3. As before, these results were found for both the Poisson and the Negbin II specification.

Another analogy between the results for P2 and P3 was found, when performing diagnostic checks: As hypothesised from the inspection of Figure 4, the Negbin II specification is more appropriate to model the time series of P3, than its Poisson counterpart. While the regression residuals of both models reveal no remaining dependence and have a mean close to zero, the residual variance in the Negbin II case is closer to one. Furthermore, the corresponding information criteria are considerably smaller than those for the Poisson model and - most striking - comparing the ratio P/df of both models leads to 1.76 versus 1.02, so the Negbin II specification provides a better fit. However, the same remark as before applies concerning model evaluation: A final decision is made upon the predictive ability.

4.2.4 Priority 4

As a last step, the estimation step was performed for P4, which is in terms of its size at best comparable to P1, with a median of only 2 emergency arrivals per day. The autocorrelogram shows slight autocorrelation at lag 3 and 4, Figure 4 depicts the highly skewed distribution of P4 which seems to be fit best by the negative binomial distribution. As shown in Table 8, regression analysis revealed Monday, Wednesday and Thursday as predictors for the time series of P4, having a positive impact compared to the rest of the week. An even stronger effect could be identified for D_{PH} and D_{afPH}, which is negative for the former and positive

Poisson model		Negbin II model	
$EA_t^{(3)}$	Coefficient	$EA_t^{(3)}$	Coefficient
$D_{WEDNESDAY}$	-.0747*	$D_{WEDNESDAY}$	-.0770*
$D_{THURSDAY}$	-.0663*	$D_{THURSDAY}$	-.0673*
D_{FRIDAY}	-.1567**	D_{FRIDAY}	-.1590**
$D_{SATURDAY}$	-.0721*	$D_{SATURDAY}$	-.0724*
D_{SUNDAY}	-.1560**	D_{SUNDAY}	-.1614**
D_{afPH}	.2751**	D_{afPH}	.2777**
$D_{TEMPMAX30}$	.1199*	$D_{TEMPMAX30}$	.1313**
$EA_{t-1}^{(3)}$	.0079**	$EA_{t-1}^{(3)}$	.0079**
$EA_{t-21}^{(3)}$	.0036**	$EA_{t-21}^{(3)}$	.0037**
$EA_{t-28}^{(3)}$	.0029**	$EA_{t-28}^{(3)}$	.0028*
constant	3.0995	constant	3.0953
$\hat{\alpha}$	-	$\hat{\alpha}$	.0216**
z_t		z_t	
mean	-.0005	mean	-.0003
std. err.	1.3163	std. err.	.9998
T_{LB}	45.1311 (0.27)	T_{LB}	44.6818 (0.28)
P/df	1.7648	P/df	1.0181
AIC	3910.232	AIC	3810.426
BIC	3957.621	BIC	3857.815

Notes: (*) indicates statistical significance on the 5%-, (**) on the 1%-level.

Table 8: Regression results for P3

for the latter predictor, respectively. Additionally, there is a negative effect on $EA_t^{(4)}$ in months defined as winter, that is October through December, so the time series exhibits some seasonality. As predicted from the inspection of the autocorrelogram, the lagged values $EA_{t-3}^{(4)}$ and $EA_{t-4}^{(4)}$ can also serve as a predictor for the present number of emergency arrivals. Albeit yielding the same outcome with regard to the size of the effects, fitting a Negbin II model once more leads to better diagnostic results: The Pearson residual and the Pearson statistic have the properties they should have in theory when the model is correctly specified, which supports the observation made in Figure 4, that the negative binomial distribution is closer to the distribution of $EA_t^{(4)}$.

Poisson model		Negbin II model	
$EA_t^{(4)}$	Coefficient	$EA_t^{(4)}$	Coefficient
D_{MONDAY}	.4616**	D_{MONDAY}	.4585**
$D_{WEDNESDAY}$	.3766**	$D_{WEDNESDAY}$	.3663**
$D_{THURSDAY}$	.2686**	$D_{THURSDAY}$	.2576**
D_{PH}	-.7283**	D_{PH}	-.7193**
D_{afPH}	.4898**	D_{afPH}	.4827*
D_{WINTER}	-.2637**	D_{WINTER}	-.2730**
$EA_{t-3}^{(4)}$	.0305*	$EA_{t-3}^{(4)}$	.0299*
$EA_{t-4}^{(4)}$	.0309*	$EA_{t-4}^{(4)}$	.0289*
constant	.7190	constant	.7317
$\hat{\alpha}$	-	$\hat{\alpha}$	.2465**
z_t		z_t	
mean	.0004	mean	.0002
std. err.	1.3109	std. err.	1.0021
T_{LB}	33.1480 (0.77)	T_{LB}	34.3368 (0.72)
P/df	1.7429	P/df	1.0185
AIC	2466.706	AIC	2369.566
BIC	2505.864	BIC	2408.724

Notes: (*) indicates statistical significance on the 5%-, (**) on the 1%-level.

Table 9: Regression results for P4

4.3 The Forecast

After completing the step of model identification, estimation and in-sample evaluation, the predictive ability of the models will be assessed in this paragraph. This will be achieved in the following way: First, one- and seven-step ahead point forecasts into the prediction sample will be constructed from each model. The important point at this stage is, that a recursive forecasting environment is used for this purpose. That means, that the models from above are only applied to construct the first forecast and are re-estimated afterwards, using one additional observation and therefore more information. The assumption underlying this procedure is, that the parameters of the models are stable over time and do not change after re-estimation due to structural breaks, which is reasonable for the present application. Additionally, a quadratic loss function as presented above will be assumed, both for mathematical convenience and for practical reasonability. With regard to the information set, large parts of the identified models are of deterministic nature and already known, when the forecast is constructed. For the climatic variables, the actual value in the predicted period $t + h$ is used to construct the forecast, assuming perfect weather forecasts and given the fact, that the influence of climatic variables in the models is rather small. Putting all these pieces together, it is possible to construct a series of point forecasts and compare them to the observed values in the prediction sample. This in turn yields a series of forecast errors, which can be assessed by applying the MSE and the RMSE. As already mentioned above, Poisson and Negbin II model yield the same point forecast for a given set of regressors and lagged dependent variables, since their conditional means are equally specified. As an extension of the point forecasts obtained from the different models, one-step and seven-step ahead density forecasts will be constructed as well by using the series of point forecasts generated before. At this point, Poisson and Negbin II models lead to different forecasts, because their probability mass functions are specified differently. The density forecasts are evaluated using the ASE afterwards.

4.3.1 Priority 1

The procedure of constructing and evaluating point forecasts is illustrated in detail for the case of one-step ahead forecasts for P1 and should serve as an example for all other cases, which involve considerably more parameters but follow the same logic. Estimation led to a Poisson and a Negbin II model for P1, where the conditional mean is specified as shown in (10). Thus, under a quadratic loss function and applying the theory from chapter three, it is straightforward to construct the optimal one-step ahead forecast for P1 as

$$f^*_{t,1} = \mathbb{E}[EA^{(1)}_{t+1}|I_t] = \mu^{P1}_{t+1|t}$$
$$= exp(1.61 + 0.15 \cdot 0 + 0.02 \cdot 7 - 0.02 \cdot 5) = 5.21, \quad (35)$$

where $I_t = (D_{MONDAY}, EA^{(1)}_{t-1}, EA^{(1)}_{t-17})$ and $e_{t,1} = 12 - 5.21 = 6.79$. Obviously, this is a poor result with a large forecast error which is due to the cause, that only three predictors for P1 could be identified in total at the estimation step, two of them being of stochastic nature. In order to generate a series of forecast errors, above forecasting procedure was repeated for the entire prediction sample. This resulted in a RMSE of 2.78 for the Poisson and the Negbin II specification, respectively, since they rely on the same conditional mean. The result indicates, that on average, the forecasts for P1 generated by the models are wrong by 2.78 arrivals per day. Using the series of one-step ahead point forecasts, a series of one-step ahead density forecasts was constructed for the entire prediction sample and for the Poisson as well as the Negbin II model. For the former, this gives an $\sqrt{ASE}$ of 0.0197, indicating, that on average, the forecast densities differ by 1.97 p.p. for a single outcome from the true densities, if they follow a Poisson distribution, which is reasonable to assume from Figure 4. For the Negbin II model, comparing the forecast densities to the true density returns $\sqrt{ASE} = 0.0192$ or an average difference of 1.92 p.p., which is slightly lower than in the case of the Poisson model.

For the seven-step ahead forecast, $f_{t,7}$, the forecasting procedure was slightly changed, because in that case, $EA^{(1)}_{t-1}$, for instance, in fact means $EA^{(1)}_{t+7-1}$, which needs to be predicted as well. For that purpose, a series of one-step ahead forecasts was generated to reach the seven-step ahead forecast in the end recursively,

Poisson model		Negbin II model	
AIC	2572.422	AIC	2565.019
BIC	2589.734	BIC	2582.331
One-step ahead forecasting performance			
RMSE	2.78	RMSE	2.78
$\sqrt{ASE}$	0.0197	$\sqrt{ASE}$	0.0192
Seven-step ahead forecasting performance			
RMSE	2.75	RMSE	2.75
$\sqrt{ASE}$	0.0206	$\sqrt{ASE}$	0.0201

Table 10: Model comparison for P1

as explained in González-Rivera (2013, pp. 187). The outcome of this procedure and the construction of a series of seven-step ahead forecast errors was a slightly lower RMSE of 2.75 for both models. Afterwards, the series of point forecasts was again used to construct the series of density forecasts, leading to $\sqrt{ASE} = 0.0206$ for the Poisson and $\sqrt{ASE} = 0.0201$ for the Negbin II model. Remarkably, both values are only slightly higher than for the one-step ahead forecasts. This might stem from the fact, that the seven-step ahead forecast corresponds to the same day of the week as the day, when the forecast is constructed.

Considering slightly smaller information criteria and $\sqrt{ASE}$ for the Negbin II specification in Table 9, this model performs best in forecasting the time series of emergency arrival counts in P1.

4.3.2 Priority 2

Turning the attention to P2, where many predictors could be identified at the estimation step, one might hypothesise, that this time series is therefore more predictable than it is the case for P1. Additionally, most of the parameters in the models are of deterministic nature, which might increase the predictive ability even further. Starting with the series of one-step ahead forecasts and the respective forecast errors, this hypothesis can be confirmed. The Poisson models lead to a RMSE of 9.48 and 9.50, both Negbin II models to RMSE= 9.49, which is a lot better compared to the average daily arrivals of P2 than it is for P1. Since the slight differences in the RMSEs are due to rounding, it is not possible to discriminate between Poisson and Negbin II model 1 or 2 at this stage. The time series of the one-step ahead forecasts (red line) for both Poisson models are plotted in Figure 5 along with the observed emergency arrivals (green line) and the corresponding 95%-confidence bands (grey area).

Figure 5: Observed arrivals and one-step ahead forecasts P2

Constructing density forecasts from the point forecasts yields an entire probability mass function for each element of the prediction sample. An example: The one-step ahead forecast for $t = 578$ using the Poisson and Negbin II model 1 is $f_{t,1} = 57.96$ and $f_{t,1} = 58.64$ for model 2, respectively, the observed value is $EA^{(2)}_{t+1} = 60$. Plugging these values into the probability mass function of a Poisson distribution returns three densities, two of them being a density forecast $f_{t,h}(Y_{t+h}|I_t)$ and one being the true density $f_{t+h}(y_{t+h}|\mathbf{x}_t, \mathbf{y}_{t+h-q})$ which are depicted in Figure 6 for illustrative purposes. On the abscissa, the number of emergency arrivals is shown, the ordinate reports the corresponding density. Note, that the discrete Poisson densities are again depicted continuously for reasons of comparability.

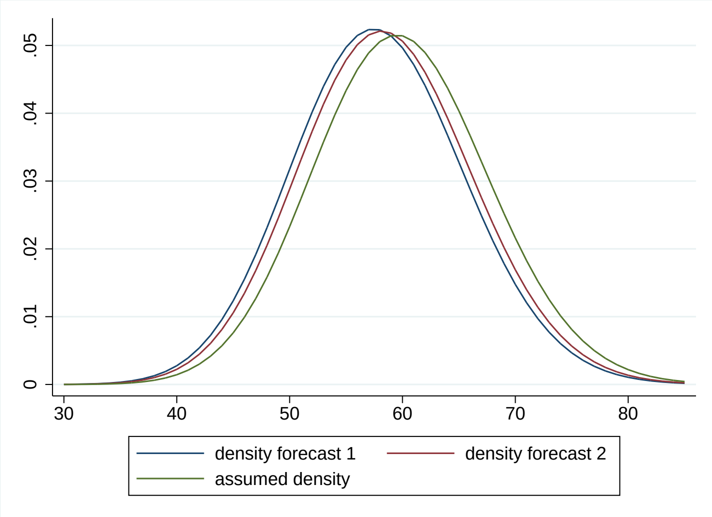

Figure 6: Poisson density forecasts for P2 (t=578)

Overall assessment of the density forecasts results in $\sqrt{ASE} = 0.0109$ for both Poisson models and $\sqrt{ASE} = 0.0108$ for the first and $\sqrt{ASE} = 0.0099$ for the second Negbin II model. Remarkably, all of the values are smaller than for the P1 density forecasts, which provides evidence for the hypothesis, that the P2 series is more predictable.

Generating the series of seven-step ahead forecasts, the same observation as in the case of P1 can be made, since the RMSE for all models slightly decreases. For Poisson and Negbin II model 1, it is RMSE = 9.28, and for model 2, RMSE = 9.30. With regard to the density forecasts, there is a slight increase in the $\sqrt{ASE}$ for all models as shown in Table 10. Again, as for the one-step ahead forecasts, the Negbin II model 2 performs best for the series of seven-step ahead density forecasts. Additionally, this specification provided the best fit and the smallest information criteria among all models for P2 and should therefore be preferred to them when forecasting this time series.

Poisson model 1		Poisson model 2	
AIC	4060.688	AIC	4055.455
BIC	4121.178	BIC	4111.625
One-step ahead forecasting performance			
RMSE	9.48	RMSE	9.50
$\sqrt{ASE}$	0.0109	$\sqrt{ASE}$	0.0109
Seven-step ahead forecasting performance			
RMSE	9.28	RMSE	9.30
$\sqrt{ASE}$	0.0110	$\sqrt{ASE}$	0.0111
Negbin II model 1		Negbin II model 2	
AIC	4017.67	AIC	4013.413
BIC	4078.161	BIC	4069.583
One-step ahead forecasting performance			
RMSE	9.49	RMSE	9.49
$\sqrt{ASE}$	0.0108	$\sqrt{ASE}$	0.0099
Seven-step ahead forecasting performance			
RMSE	9.28	RMSE	9.30
$\sqrt{ASE}$	0.0111	$\sqrt{ASE}$	0.0103

Table 11: Model comparison for P2

4.3.3 Priority 3

After illustrating the forecasting procedure and its results using various examples in the previous sections, the results for P3 and P4 are illustrated more briefly. Adopting the hypothesis made concerning P2, one could assume that the time series of P3 is also well predictable, since several predictors could be identified at the estimation step. Additionally, both evidence from the diagnostic results in Table 7 and Figure 4 suggest, that a Negbin II specification might be superior to a Poisson specification in forecasting the emergency arrivals of P3. For the series of one-step ahead forecasts, the models return a RMSE of 6.401, which equals $(6.401/33.94) \cdot 100 = 18.86\%$ of the time series' mean. With regard to the one-step ahead density forecasts, the Poisson model yields $\sqrt{ASE} = 0.0112$, the Negbin II model $\sqrt{ASE} = 0.0095$, that is, it differs by 7 p.p. less from the true density on average, which is in agreement with the hypothesis made above.

For the seven-step ahead forecasts, the situation is the same: The models account for a RMSE of 6.1887, which is again less than for the one-step ahead forecasts and for the forecast densities, the Poisson model yields $\sqrt{ASE} = 0.0104$

and the Negbin II model $\sqrt{ASE} = 0.0087$, respectively. Summing up the results in Table 11, the Negbin II specification turns out to be the best model to forecast the time series of P3.

Poisson model		Negbin II model	
AIC	3910.232	AIC	3810.426
BIC	3957.621	BIC	3857.815
One-step ahead forecasting performance			
RMSE	6.401	RMSE	6.401
$\sqrt{ASE}$	0.0112	$\sqrt{ASE}$	0.0095
Seven-step ahead forecasting performance			
RMSE	6.1887	RMSE	6.1887
$\sqrt{ASE}$	0.0104	$\sqrt{ASE}$	0.0087

Table 12: Model comparison for P3

4.3.4 Priority 4

By continuing the line of argumentation, which was applied to the three previous priority groups, the time series of P4 should be less predictable than P2 and P3 and a Negbin II specification might result in better predictive ability than a Poisson model, because the distribution of P4 emergency arrivals is highly skewed. In general, one might hypothesise that the results for P4 are therefore more similar to P1, than to P2 and P3. And in fact, this is the case: For the series of one-step ahead forecasts, the models show a RMSE of 2.4457, which is considerably less than the RMSEs for P2 and P3, but accounts for $(2.4457/2.84) \cdot 100 = 86.12\%$ of the time series' overall mean. This turns out to be a large deviation, indicating poor predictability, even though strong deterministic predictors could be identified when estimating the models. This result might be partly caused by the fact, that the underlying distribution of emergency arrivals is the most skewed among all priority groups (see Table 1 and Figure 4), which implies a higher probability mass in the tails of the distribution. This, however, is not taken into account when computing the point forecasts, since they only represent future outcomes of the conditional mean. One is inclined to infer, that the construction of density forecasts is especially fruitful for this particular situation. And they lead to better results, indeed: For the Poisson model, it is $\sqrt{ASE} = 0.0189$, for the Negbin II model $\sqrt{ASE} = 0.0131$, indicating a deviation, which is comparable to the size found for the other priority groups and providing evidence in favour of the hypothesis, that a Negbin II specification might be preferable.

For the series of seven-step ahead forecasts, the situation does not change. Again, both models yield a RMSE of 2.6072, which is huge compared to the time series mean, but lead to $\sqrt{ASE} = 0.0199$ for the Poisson, and $\sqrt{ASE} = 0.0136$ for the Negbin II model. The summary of in-sample and out-of-sample evaluation criteria shows, that for P4, just as for all previous priority groups, the Negbin II model serves best in forecasting the time series of emergency patient arrival counts.

Poisson model		Negbin II model	
AIC	2466.706	AIC	2369.566
BIC	2505.864	BIC	2408.724
One-step ahead forecasting performance			
RMSE	2.4457	RMSE	2.4457
$\sqrt{ASE}$	0.0189	$\sqrt{ASE}$	0.0131
Seven-step ahead forecasting performance			
RMSE	2.6072	RMSE	2.4938
$\sqrt{ASE}$	0.0199	$\sqrt{ASE}$	0.0136

Table 13: Model comparison for P4

5 Conclusion

The present paper made the attempt to forecast daily emergency patient arrival counts classified by priority groups using appropriate statistical techniques. For that purpose, regression models based on discrete probability distribution where presented and applied to data from a German hospital. After estimating the model parameters, they were used to construct point and density forecasts for each priority group separately, both for one and for seven days ahead. This last paragraph critically summarises what could be achieved and which parts remain unsolved or beyond the scope of this examination.

From a statistical point of view, one important insight is, that the distributions of emergency patient arrivals are indeed well approximated by discrete probability distributions such as the Poisson and the negative binomial, which makes it reasonable to depart from ARIMA modelling and to apply appropriate models for count data. These models led to point forecasts, which are relatively sound, at least for P2 and P3. Their main advantage, however, lies in the construction of density forecasts, since both Poisson and Negbin II models yielded density forecasts close to the assumed true density and with better global performance than the series of point forecasts. This is of particular importance in the presence of skewed distributions, where reporting a future conditional mean as a point forecast is even less meaningful than it is in the case of the normal distribution. A further purely statistical result is the fact, that the Negbin II specification provides better diagnostic results after estimating the model and better forecasting performance for all priority groups. This suggests, that forecasting the series of emergency arrival counts should be performed using the Negbin II models.

With regard to the predictors, which could be identified when estimating the different models, the results are in large part in agreement with already existing results in the literature. This observation applies to the finding, that in general the day of the week and the specific nature of a day are the strongest predictor for the number of emergency arrivals. The typical behaviour with a peak on Mondays, followed by a decrease towards the weekend could be confirmed, at least for some of the priority groups. A yearly seasonality could be identified as well, being, however, less pronounced than the day of the week and contrary to Batal et al. (2001) among others, who find a peak during the winter. Additionally, the role of climatic factors influencing the number of emergency arrivals remains ambiguous, as outlined in chapter two, since no distinct result could be identified. Contrary to the research conducted by McCarthy et al. (2008), the different time series revealed a pronounced dependence structure, both in the raw data and in estimated autoregressive coefficients, which are statistically different from zero in a significant way. This might be an indication of heterogeneity with regard to the level of aggregation, since McCarthy et al. (2008) base their investigation on hourly data. Another indication of heterogeneity, which is in complete agreement with the results obtained by Sun et al. (2009), is related to the different priority groups. It is one main result of this investigation, that the groups differ remarkably in size, influencing factors and predictability. Most strikingly, small groups like P1 and P4 turned out to be less predictable. For P1, the group of the highest acuity, this is exactly the same result as presented by Sun et al. (2009). The conclusion drawn from this finding is, that emergency arrivals should be forecast separately, classified by severity of the illness or an appropriate proxy-variable, as it was the case in this investigation.

Albeit finding several insightful results, it has to be stressed that they are subject to various limitations: The dataset contains observations collected over a period of less than two years, which challenges the assumption of temporal stability of the model parameters. Some of the parameters moreover depend on the explicit geographic location of the hospital providing the data, which especially concerns the effects of climatic variables and seasonality. In addition, emergency arrivals were aggregated on a daily basis and the hospital-specific classification was used to discriminate among priority groups.

As pointed out in the introduction, this study is of some practical relevance, since it might facilitate planning processes in hospitals and thus enhance efficiency. Hence, it is the question, what the practical implications of the insights illustrated above could be. First and foremost, the main implication is, that collecting data is a valuable procedure, which is indispensable in order to develop sophisticated forecasting models. From the results above, it becomes evident, that this data should be as detailed as possible, because this uncovers inherent heterogeneity and allows to forecast different patient groups separately, using the Negbin II model as suggested above.

The issue of data and its collection is also one of the elements lying beyond the scope of this single investigation. Obviously, the only factors used in the estimation of the forecasting models were those already identified in the literature. One further step would be, to include additional data on factors, which might drive the number of emergency arrivals as well, such as demographic data or data on public holidays which typically involve a higher consumption of alcohol. Other promising extensions of this study exist with regard to statistical methods. Admittedly, the results of this paper rely on the most basic versions of count data models, namely a Poisson and Negbin II dynamic regression model. The application of more sophisticated versions of these models might lead to even higher predictive ability. Another extension could be a more appropriate construction of point forecasts following Freeland and McCabe (2004), who suggest computing them using the conditional median rather than the conditional mean for discrete distributions. Additionally, there exist more sophisticated methods for evaluating density forecasts as well, especially the method of the *probability integral transform* developed by Diebold et al. (1998) which has been adjusted for the case of discrete probability distributions by Liesenfeld et al. (2008). This method could provide an assessment of density forecasts in greater detail.

To sum up the findings of this investigation, it can be said, that the combination of count data models and density forecasts can provide a powerful tool for the purpose of forecasting emergency patient arrival counts. Especially the more flexible Negbin II specification performs well in producing one- and seven-step ahead density forecasts. Since emergency arrivals are heterogeneous in many regards, it is sensible to forecast them separately, classified by the severity of their illness.

References

Batal, H., Tench, J., McMillan, S., Adams, J., and Mehler, P. S. (2001). Predicting Patient Visits to an Urgent Care Clinic Using Calendar Variables. *Academic Emergency Medicine*, 8(1):48–53.

Cameron, A. C. and Trivedi, P. K. (1986). Econometric Models Based on Count Data. Comparisons and Applications of Some Estimators and Tests. *Journal of Applied Econometrics*, 1(1):29–53.

Cameron, A. C. and Trivedi, P. K. (2013). *Regression Analysis of Count Data*, *Econometric Society Monographs*, volume 53. Cambridge University Press, 2nd edition.

Cardinal, M., Roy, R., and Lambert, J. (1999). On the Application of Integer-valued Time Series Models for the Analysis of Disease Incidence. *Statistics in Medicine*, 18(15):2025–2039.

Davis, R. A., Dunsmuir, W. T. M., and Streett, S. B. (2003). Observation-driven Models for Poisson Counts. *Biometrika*, 90(4):777–790.

Deutscher Wetterdienst Datenserver Klimadaten (2015). `http://www.dwd.de/bvbw/appmanager/bvbw/dwdwwwDesktop?_nfpb=true&_pageLabel=_dwdwww_klima_umwelt_klimadaten_deutschland&T82002gsbDocumentPath=Navigation%2FOeffentlichkeit%2FKlima__Umwelt%2FKlimadaten%2Fkldaten__kostenfrei%2Fausgabe__tageswerte__node.html%3F__nnn%3Dtrue` (online, accessed 07.09.2015).

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review*, 39(4):863–883.

Dougherty, C. (2011). *Introduction to Econometrics.* Oxford University Press, Oxford and New York, 4th edition.

Freeland, R. K. and McCabe, B. (2004). Forecasting Discrete Valued Low Count Time Series. *International Journal of Forecasting*, 20(3):427–434.

Georgii, H.-O. (2015). *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik.* De Gruyter Studium. De Gruyter, Berlin, 5th edition.

Gesellschaft Deutscher Krankenhaustag (11/14/2012). 35. Deutscher Krankenhaustag im Rahmen der MEDICA eröffnet.

González-Rivera, G. (2013). *Forecasting for Economics and Business.* The Pearson Series in Economics. Pearson, Boston.

Gourieroux, C., Monfort, A., and Trognon, A. (1984). Pseudo Maximum Likelihood Methods: Applications to Poisson Models. *Econometrica*, 52(3):701–720.

Graupe, S. (2012). Die verborgenen Quellen des Marktgehorsams — und wie wir sie verändern lernen. *Vierteljahresschrift zu wissenschaftlichen Pädagogik*, 88(4)639–658.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton, N.J.

Jones, S. A., Joy, M. P., and Pearson, J. (2002). Forecasting Demand of Emergency Care. *Health Care Management Science*, 5(4):297–305.

Jung, R. C., Kukuk, M., and Liesenfeld, R. (2006). Time Series of Count Data: Modeling, Estimation and Diagnostics. *Computational Statistics & Data Analysis*, 51(4):2350–2364.

Kam, H. J., Sung, J. O., and Park, R. W. (2010). Prediction of Daily Patient Numbers for a Regional Emergency Medical Center Using Time Series Analysis. *Healthcare Informatics Research*, 16(3):158–165.

Kirchgässner, G., Wolters, J., and Hassler, U. (2013). *Introduction to Modern Time Series Analysis*. Springer Texts in Business and Economics. Springer, Berlin and New York, 2nd edition.

Liesenfeld, R., Nolte, I., and Pohlmeier, W. (2008). Modelling Financial Transaction Price Movements: a Dynamic Integer Count Data Model. In Bauwens, L., Pohlmeier, W., and Veredas, D., eds., *High Frequency Financial Econometrics*, pages 167–197. Springer, New York.

Marcilio, I., Hajat, S., Gouveia, N., and Merchant, R. C. (2013). Forecasting Daily Emergency Department Visits Using Calendar Variables and Ambient Temperature Readings. *Academic Emergency Medicine*, 20(8):769–777.

Martin, V., Hurn, S., and Harris, D. (2013). *Econometric Modelling with Time Series: Specification, Estimation and Testing*. Themes in Modern Econometrics. Cambridge University Press, Cambridge.

McCarthy, M. L., Zeger, S. L., Ding, R., Aronsky, D., Hoot, N. R., and Kelen, G. D. (2008). The Challenge of Predicting Demand for Emergency Department Services. *Academic Emergency Medicine : Official Journal of the Society for Academic Emergency Medicine*, 15(4):337–346.

Schumpeter, J. (1933). The Common Sense of Econometrics. *Econometrica*, 1(1):5–12.

StataCorp. 2013. *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP.

Sun, Y., Heng, B., Seow, Y., and Seow, E. (2009). Forecasting Daily Attendances at an Emergency Department to Aid Resource Planning. *BMC Emergency Medicine*, 9(1):1–9.

Tandberg, D. and Qualls, C. (1994). Time Series Forecasts of Emergency Department Patient Volume, Length of Stay, and Acuity. *Annals of Emergency Medicine*, 23(2):299–306.

Tay, A. S. and Wallis, K. F. (2000). Density Forecasting: a Survey. *Journal of Forecasting*, 19(4):235–254.

Tenreiro, C. (1998). Asymptotic Distribution for a Discrete Version of Integrated Square Error of Multivariate Density Kernel Estimators. *Journal of Statistical Planning and Inference*, 69(1):133–151.

Wegman, E. J. (1972). Nonparametric Probability Density Estimation. *Journal of Statistical Computation and Simulation*, 1(3):225–245.

Winkelmann, R. (2008). *Econometric Analysis of Count Data*. Springer, Berlin, 5th edition.

Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach*. South-Western Cengage Learning, Mason, OH, 5th edition.

A On the Interpretation of Regression Coefficients

Since the conditional mean of the Poisson and the Negbin II model is specified in an exponential form, the interpretation of regression coefficients differs from the standard interpretation, which applies to the linear regression model. This non-standard interpretation is explained briefly using a simple example presented in Winkelmann (2008, p. 70) and is also valid for the more complex specifications used throughout study.

To calculate the effect of a certain regressor x_j on the conditional mean $\mathbb{E}[y_t|\mathbf{x}_j]$, one computes the partial derivative of the conditional mean with respect to the regressor, which yields

$$\frac{\partial \mathbb{E}[y_t|\mathbf{x}_j]}{\partial x_j} = exp(\mathbf{x}_j'\boldsymbol{\beta})\beta_j, \quad j = 1, \ldots, j \tag{A.1}$$

and therefore depends on $exp(\mathbf{x}_j'\boldsymbol{\beta})$, which differs across observations. In order to obtain a measure for a constant effect, one could instead consider the relative change in the conditional mean, which is given by

$$\frac{\partial \mathbb{E}[y_t|\mathbf{x}_j]/\mathbb{E}[y_t|\mathbf{x}_j]}{\partial x_j} = \beta_j. \tag{A.2}$$

The interpretation of the regression coefficient is therefore the relative change in the conditional mean due to a marginal change in the regressor.

A problem arises for models including dummy variables, since they cannot change on the margin. As Winkelmann (2008, p. 71) points out, the relative change due to variation in a dummy variable can be approximated using a linear Taylor series expansion which is valid for small regression coefficients. The approximation returns β_j as the relative change in the conditional mean due to a change in a dummy variable.

B Forecasting the Time Series of Overall Emergency Arrivals

Although it was the explicit aim of this paper, to model and forecast the different priority groups separately, the same procedure was also applied to the overall time series of emergency arrivals. The results are presented below, the estimation results first, the model evaluation based on the predictive ability afterwards. The overall time series is defined as P0, the emergency arrivals for a specific day are denoted as $EA_t^{(0)}$.

Poisson model		Negbin II model	
$EA_t^{(0)}$	Coefficient	$EA_t^{(0)}$	Coefficient
D_{MONDAY}	.1232728**	D_{MONDAY}	.1236169**
D_{SUNDAY}	-.0504205**	D_{SUNDAY}	-.0511887**
D_{afPH}	.2206338 **	D_{afPH}	.2200591**
$D_{TEMPMAX20}$	.0621778**	$D_{TEMPMAX20}$	.0627979**
$EA_{t-1}^{(0)}$	.0019769**	$EA_{t-1}^{(0)}$	.0019733**
$EA_{t-7}^{(0)}$	.0012886**	$EA_{t-7}^{(0)}$	.0012801**
constant	4.20878	constant	4.209793
z_t		z_t	
mean	-.0000643	mean	-.0000494
std. err.	1.206154	std. err.	1.001857
T_{LB}	45.7893 (0.24)	T_{LB}	36.3363 (0.64)
P/df	1.470311	P/df	1.014415
AIC	4478.952	AIC	4433.712
BIC	4509.371	BIC	4464.131

Notes: (*) indicates statistical significance on the 5%-, (**) on the 1%-level.

Table 14: Regression results for P0

Poisson model		Negbin II model	
AIC	4478.952	AIC	4433.712
BIC	4509.371	BIC	4464.131
One-step ahead forecasting performance			
RMSE	11.47	RMSE	11.47
$\sqrt{A\bar{S}E}$	0.0094	$\sqrt{A\bar{S}E}$	err.
Seven-step ahead forecasting performance			
RMSE	11.64	RMSE	11.64
$\sqrt{A\bar{S}E}$	0.0097	$\sqrt{A\bar{S}E}$	err.

Table 15: Model comparison for P0