BZPH

Berliner Zentrum Public Health

00-02

United Kingdom:

Exploratory Analysis of the

Mortality-Unemployment Relationship

with Poisson Distributed Lags

Günter Edenharter, Dieter Friedrich,
M. Harvey Brenner

Veröffentlichungsreihe des
Berliner Zentrums Public Health
ISSN 0949-0752

Veröffentlichungsreihe
des
Berliner Zentrums Public Health
ISSN 0949-0752

00-02

United Kingdom:

Exploratory Analysis of the

Mortality-Unemployment Relationship

with Poisson Distributed Lags

Günter Edenharter, Dieter Friedrich,
M. Harvey Brenner

Berlin, October 2000

Berliner Zentrum Public Health
Geschäftsstelle
Ernst-Reuter-Platz 7
10578 Berlin
Tel.: (030) 314 - 21 970
Fax: (030) 314 - 21 578

Contents

List of Figures

Summary

Distributed lag models are statistical tools that are designed to analyze the effect of non-contemporaneous associations between time series. An important example that illustrates the need for such models in epidemilogical research is the effect of smoking on the actual health status of individuals: the contemporaneous effect of smoking on health will be negligible, it is the long-term exposure over 20 - 30 years, which harms not only the bronchial system but also causes the cardiovascular system to deterioate.

First we give an introduction to the theoretical background of distributed lag modeling, describe some classes of distributed lag models and outline some of the associated model estimation and specification problems. The Poisson distributed lags are introduced as a very flexible but parameter parsimonous class of distributed lag models that need less a priori information than Polynomial distributed lag or Shiller distributed lag models. Poisson lag distributions can be summarized by single parameter which corresponds to the mean lag and the variance of the lag distribution. Their smooth functional form enhances graphic presentations and comparisons. One disadvantage of Poisson distributed lag estimators is the computational burden of non-linear optimization techniques. For our analysis we used a general purpose simulated annealing algorithm.

Five time series from the United Kingdom are used as an exploratory example for epidemiological distributed lag modeling based on Poisson lags: Total mortality rate as dependent variable and unemployment rate, spirits consumption, cigarette consumption and real gross domestic product as independent variables. Data for the five time series were available for years 1950 - 1994.

The time series are plotted for levels and first differences in conjunction with the associated autocorrelation functions. Three of the time series, i.e., unemployment rate, spirits and cigarette consumption exhibit a structural break within the level series, the break-point year can be estimated roughly for all the three series at about 1970. With the exception of total mortality, all the time series exhibit heteroscedasticity in first differences.

The comparison of the single indendent variable models (Model 1-4: levels, Model 5-8: first differences) with the simultanousely estimated four indendent variable model 9 (levels) and model 10 (first differences) reveal dramatic differences for the corresponding 'crude' and 'adjusted' estimated Poisson lag distributions, e.g. the mean lag for cigarette consumption increased from 5.80 years (model 3) to circa 10 years (model 9).

The results demonstrate the importance of appropriate adjustment in epidemiological aggregated time series regression models and encourage the further development and refinement of econometric tools for applications in epidemiological research.

1 Distributed lag modeling

1.1 Theoretical background

Very often the effect of an event persists over time so that for the measurement of an overall effect it is not sufficient to analyse contemporaneous associations between two variables. In many cases the magnitude of an actual value of a dependent variable can only be explained when the impact of earlier values of a independent variables have been taken into account. An important example from epidemiological research is the effect of smoking on the actual health status of individuals: the contemporaneous effect of smoking on health will be negligible, it is the long-term exposure over 20 - 30 years, which harms not only the bronchial system but also causes the cardiovascular system to deterioate. Our interest here is, to provide a methodological framework for the analysis of the development of mortality levels, that have experienced significant long-term declines around the world (Murray/Chen [1993]).

To derive a simple statistical model that allows us the analysis of non-contemporaneous relationships between variables, we start from the classical linear regression model with a dependent variable y_t and only one independent variable x_t:

$$y_t = \beta_0 + \beta_1 x_t + u_t \quad t = 1, \ldots, T. \tag{1}$$

To keep things simple we assume that u_t is a zero-mean white-noise process with variance σ_u^2. A straightforward approach to take into account the effect of earlier values of the independent variable on y_t is to extend equation (1) as follows:

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \ldots + u_t \quad t = 1, \ldots, T, \tag{2}$$

i.e., we add lagged values of the independent variables x_t.

Equation (2) can be rewritten as:

$$y_t = \alpha + \sum_{i=0}^{\infty} \beta_i x_{t-i} + u_t \quad t = 1, \ldots, T, \tag{3}$$

The general form of equation (3) allows for an infinite lag length. In most situations there exist a number $M < \infty$ with $\beta_i = 0$ for $i > M$. This number M is called the maximum lag length. With a finite lag length equation (3) can be written as:

$$y_t = \alpha + \sum_{i=0}^{M} \beta_i x_{t-i} + u_t \quad t = 1, \ldots, T. \tag{4}$$

Alternative methods and their effects for the treatment of the truncation remainder

$$R = \sum_{i=M+1}^{\infty} \beta_i x_{t-i}$$

are discussed by Schmidt [1976] and Schmidt/Guilkey [1976].
It is possible to reparameterize equation (4) as

$$y_t = \alpha + \beta \sum_{i=0}^{M} w_i x_{t-i} + u_t \quad t = 1, \ldots, T, \tag{5}$$

with

$$w_i = \frac{\beta_i}{\sum_0^M \beta_i} \qquad \text{and} \qquad \sum_0^M w_i = 1.$$

We will refer to the lag weights $(w_0, \ldots, w_M)$ as the normalized lag weights or the normalized lag distribution. A useful statistic by which to compare different lag distributions is the mean lag $\bar{w}$:

$$\bar{w} = \sum_{i=0}^{M} i w_i,$$

provided that the normalized lag weights w_i have equal sign and the lag distribution is unimodal.
The naive approach, to try to estimate the parameters $\alpha, \beta_0, \ldots, \beta_q$ from equation (4) with the usual Ordinary Least Squares (OLS) approach, will run into at least two difficulties:

1. The maximum lag length M has to be determined. For all practical purposes M has to be so much smaller than the number of available observations that the remaining number of observations, $T - M$, is large enough to allow for useful statistical inference from the regression analysis.

2. The lagged independent variables $x_{t-1}, \ldots, x_{t-M}$ will show a high degree of multicollinearity, i.e., nearly linear dependence. This may make it numerically difficult to solve the standard normal equations from the OLS approach. Even when a numerical solution is found, which is the normal case with modern computer hardware, the solutions are unstable in the sense that small changes in the model, e.g. increasing or decreasing the maximum lag length M by one, may result in dramatic changes of the estimated lag weight coefficents $(b_0, \ldots, b_M)$.

The determination of the maximum lag length M can be done simply by educated guessing, testing procedures or by the use of model-order selection criteria, such as AIC (Akaike [1974]) or BIC (Schwarz [1978]). As there is no space here to provide even an elementary discussion of the pros and cons of different approaches for the determination of the maximum lag length, we assume for the remaining part of this paper that the maximum lag length is theoretically given.

To overcome the problem of multicollinearity various solutions have been proposed which have one feature in common (see Judge et al. [1985, p. 355ff]): All these solutions put restrictions on the coefficients $\beta_0 \ldots, \beta_M$, i.e., the dimension of the parameter space is reduced by assuming that the lag weights β_i are values of a function $f(i)$.

As early as 1937, Fisher[1937] proposed arithmetically declining lags, with the function $f(i)$ given by:

$$\beta_i = f(i) = \begin{cases} (M+1-i)c & i = 0, \ldots, M \\ 0 & i > M \end{cases} \tag{6}$$

An appropriate parameter c has to be provided by the user.

The approach of Fisher as given in equation (6) was a computationally simple solution, but his approach was very specific and inflexible.

Circa 30 years later a far more general approach to overcome the multicollinearity problem has become very popular among econometricians: the Polynomial distributed lag (PDL) or Almon distributed lag when it is named after the author (Almon[1965]). She suggested that the function $f(i)$ should be a polynomial of degree $q < M$, i.e.

$$\beta_i = c_0 + c_1 i + c_2 i^2 + \ldots c_q i^q. \tag{7}$$

The polynomial lag shape is very flexible and the parameters could be estimated easily. The presentation of the Almon distributed lag scheme as a special Restricted Least Squares (RSL) problem made it easier to analyse the statistical properties of the Almon lag estimator. With the RSL presentation, a polynomial of degree q may be imposed on the lag weights by specifying $M - q$ linear homogenous restrictions of the form

$$\boldsymbol{R\beta}=\mathbf{0} \tag{8}$$

on the model $\boldsymbol{y}=\boldsymbol{X\beta}+\boldsymbol{u}$. The elements of the restricting matrix $\boldsymbol{R}$ can be derived from the well known Pascal´s triangle. It was also easy to add additional constraints on single

lag weights or so called endpoint constraints with $\beta_{-1} = 0$ and $\beta_{M+1} = 0$, although these endpoint constraints could have unintended side effects.
The Almon lag scheme was more flexible than other lag schemes, but Shiller[1973] thought that it may be too restrictive since the functional form of the lag weights is supposed to follow exactly a polynomial of order q. His solution imposed not a deterministic constraint on the lag weights as given in equation (8) but a stochastic constraint:

$$\boldsymbol{R\beta}=\boldsymbol{v} \tag{9}$$

It is assumed that the stochastic vector $\boldsymbol{v}$ has a skalar variance-covariance matrix of the form $E(\boldsymbol{vv}') = (\sigma^2/k^2)\boldsymbol{I}$. With an a priori specified smoothing parameter k the Shiller restricted lag weight estimator $\hat{\boldsymbol{\beta}}= (\hat{\beta}_0, \ldots, \hat{\beta_M})'$ can be simply derived as an extension to the solution of the OLS normal equations:

$$\hat{\boldsymbol{\beta}} = [\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{R}'\boldsymbol{R}]^{-1}\boldsymbol{X}'\boldsymbol{y}. \tag{10}$$

The smoothing parameter k controls the form of the lag weights: for $k = 0$ the solution of the Shiller restricted lag weight estimators is identical to the unrestricted OLS solution, for $k \to \infty$ the solution approaches the Almon lag estimator.
Since the numerical algorithms for the computation of Almon lag estimators and Shiller lag estimators are easily implemented as the solutions of special linear equation systems, these two approaches are the most common solutions which are provided by widespread commercial econometric software packages like RATS$^{(\mathrm{TM})}$.
Even when the greater flexibility of the Shiller lag estimator provides an advantage over the Almon lag estimator, the price to pay for this comfort is that not only an appropriate polynomial order q has to be determined, but also an appropriate smoothing parameter k has to be found. Shiller[1973] suggested a 'rule of thumb' for the determination of the smoothing parameter k but the statistical validity for such a procedure is questionable. It has to be emphasized that when maximum lag length, polynomial order and the smoothing parameter are determined by an iterative 'data squeezing' process, the statistical properties of the finally resulting estimators are completely undefined so that any valid statistical inference is impossible.

1.2 Poisson distributed lags

One approach to overcome the problem of insufficient a priori information is to use another class of distributed lag estimators, which provides enough flexibility with regard to

functional form but minimizes the requirements for additional a priori parameters. Here we will introduce the Poisson distributed lag model (Friedrich[1982]).
Poisson lags are derived directly from the Poisson probability distribution, and the normalized lag weights are given by:

$$w_i(\lambda) = e^{-\lambda}\frac{\lambda^i}{i!} \qquad i \geq 0. \tag{11}$$

The mean lag and the variance of the Poisson lag distribution as specified by equation (11) is identical to the parameter λ, i.e. $\bar{w} = \lambda$. In dependence of only one parameter λ we get very flexible lag distributions, typical shapes for Poisson lag distributions are given in Fig. 1.2 (a)-(c) . The shape of the poisson lag distributions is very different for $\lambda < 1$ in comparison to $\lambda > 1$. With $\lambda < 1$ the Poisson lag distribution has a maximum at lag $i = 0$ and the values decline very fast to zero (Fig. 1.2 (a)). With $\lambda > 1$ the lag distribution first increases, reaches a maximum and then declines, in dependence of λ the increase and decline may be asymmetric ($\lambda = 1.5, 3.0$) or more or less symmetric ($\lambda = 7.0$).
In practical situations, especially with a limited number of observations and a too small data-enforced maximum lag length M, another shape of the Poisson lag distribution may be seen which is described in Fig. 1.2 (c). These Poisson lag distributions are right truncated and may even have their global maximum at the maximum lag length M. In the case of truncated Poisson lag distributions it is no longer valid that the mean lag is identical to parameter λ. For this reason we calculate the mean lag for the Poisson lag distribution as usual, i.e.:

$$\bar{\lambda} = \sum_{i=0}^{M} i w_i(\lambda),$$

where the slight change in notation emphasizes just the fact that the mean lag is calculated from a Poisson lag distribution.
In econometrics Poisson lag distributions have a special advantage as this class of lag distributions can be theoretically derived from basic behavioral reaction equations for adjustment processes (Friedrich[1982]). Other examples for the use of economic reasoning for the derivation of special lag distributions can be found in Grilliches[1968] or Nerlove[1972]. Up to now, no analogue approach is used as foundation of epidemiological distributed lag modeling.
The Poisson lag model can be easily extended to the case of several independent variables:

$$y_t = \alpha + \sum_{k=1}^{K} \beta_k \sum_{i=0}^{M_k} e^{-\lambda_k}\frac{\lambda_k^i}{i!} x_{t-i,k} + u_t \qquad t = 1, \ldots, T. \tag{12}$$

In the case of K independent variables the use of the poisson lag distribution has the advantage that the number of parameters to be estimated can be reduced to an absolute minimum of $2K + 1$. This results in a maximum number $T - 2K - 1$ for the degrees of freedom which makes the poisson distributed lag model a natural approach for distributed lag modeling with short time series.
As for several independent variables, the Poisson lag distributions can be estimated simultaneously, this approach is especially suited to make comparisons between ‘crude’ estimators and ‘adjusted’ estimators when we refer to the terms from epidemiology: With a crude estimator the relationship between two variables is analyzed, with an adjusted estimator the change of a crude estimator is analyzed when additional, possibly confounding, variables are included in a model, and all parameters are estimated simultaneously. Such adjusted estimators are extremely important, as bivariate statistical analysis can be very misleading. This can also be demonstrated with the results that we have found in our study.
A closer look at equation (12) reveals a problem which also explains why the use of Poisson lag distributions is not as widespread as the use of Almon lags or Shiller lags: The parameters $\lambda_1, \ldots, \lambda_k$ enter into the equation in a non-linear manner, a solution to the associated non-linear least squares problem cannot be calculated by solving a simple system of linear equations. When estimators for the parameters $\lambda_1, \ldots, \lambda_K$ have been obtained by non-linear optimization techniques, the estimators of $\beta_1, \ldots, \beta_K$ can be calculated via the usual OLS approach.
The discussion of non-linear optimization techniques is far beyond the scope of this short introduction to distributed lag modeling. After a lengthy search process for a reliable procedure we decided to use a modern stochastic optimization technique, simulated annealing, for the calculation of estimators for the parameters $\lambda_1, \ldots, \lambda_k$. We summarize briefly the concept and the advantages of the simulated annealing approach for non-linear optimization (Goffe et al. [1994]):
Simulated annealing´s roots are in thermodynamics, where one studies a system´s thermal energy. A description of the cooling of molten metal motivates this algorithm. After slow cooling (annealing), the metal arrives at a low energy state. Inherent random fluctuations in energy allows the annealing system to escape local energy minima to achieve the global minimum. But if cooled very quickly (or ’quenched’), it might not escape local energy minima and when fully cooled it my contain more energy than annealed metal. Simulated annealing attempts to minimize some analogue of energy in a manner similar to annealing to find the global optimum. One important advantage of simulated annealing is that it

can escape from local maxima or minima and can maximize or minimize functions that are otherwise difficult or impossible to optimize. One special advantage for the estimation of the parameters $\lambda_1, \ldots, \lambda_k$ was that the restrictions $0 \leq \lambda_k \leq M_k$ could be easily enforced, which was not possible with the other non-linear optimization algorithms that had been tested. Another advantage is that when the stochastic optimization process of the simulating annealing algorithm is run twice for the same function, with the initial seed of the random number generator set to different values, and the two found optima are identical, then we can have considerable confidence that this optimum is really the global optimum. We used the GAUSS implementation of Bill Goffe´s simulated annealing program which was written by E.G. Tsionas[1995] from the Department of Economics at the University of Toronto. One, and probably the only, disadvantage of the simulated annealing algorithm is that the many functions evaluations that are necessary to find a global optimum require a good deal of computational power. This disadvantage is increasingly vanishing as computers seem to become much more powerful from year to year, but actually this disadvantage is large enough not to use it as a daily routine algorithm for the computation of distributed lag models.
Remark:
When we compare different Poisson lag distributions we plot the normalized lag weights $w_{i,k}$ and print the associated mean lag $\bar{\lambda}_k$ nearby. Please note that the evaluation of the total effect also requires taking into account the associated parameter β_k, which is especially important because of the sign of β_k.

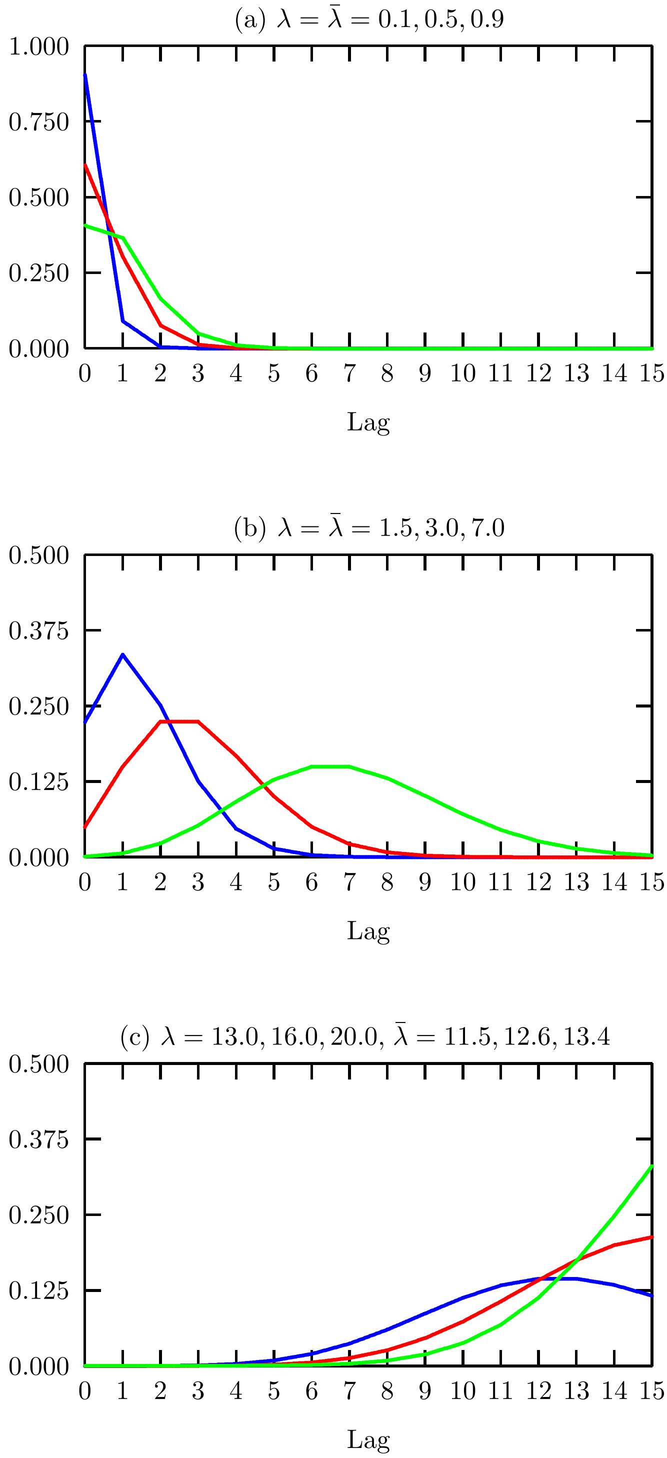


Figure 1: Typical shapes of Poission lag distributions

2 Results

2.1 Time series properties of model variables

With Figs. 2 to 6 we provide a summary of the most important time series characteristics for all the variables that we will use in our exploratory Poisson distributed lag modeling. Data for the five time series were available for years 1950 - 1994.

1. **Total mortality rate** (RTMORT), Fig. 2. As the dependent variable we use the age- and sex- adjusted total mortality rate. The reference population which was used for the direct standardization is given in appendix A. The level of the total mortality rate shows a remarkable secular decline from 1950 to 1994, which is typical not only for for all the EU countries and the US which were analysed in our study but for all industrialized nations outside of Eastern Europe. The shape of the associated autocorrelation function is typical for a non-stationary time series: the autocorrelations are very high and decay only slowly with increasing lag length, for lag 11 and higher all the autocorrelation values are within the approxmative 95% confidence intervals which are plotted in all autocorrelation plots with two small dashed lines. When we look at the total mortality rate in first differences no specific trend pattern can be recognized. The time series fluctuates irregularly around a constant. The associated autocorrelation function for the first differences of the total mortality rate decays very quickly. There is only one value of the autocorrelation at lag 1 which is not within the 95% confidence interval. From a time series perspective we may characterize the total mortality rate as a typical random walk with negative drift, although such a time series with a stochastic trend may be difficult to distinguish from a time series with a deterministic trend. Even without using statistical inference procedures, such as unit root tests (Banerjee et al. [1993]), we can infer from the basic descriptive analysis that the total mortality rate is a stationary time series in first differences or, in other words, the total mortality rate is integrated of order one, i.e., $I(1)$.

2. **Unemployment Rate** (UR), Fig. 3. The unemployment rate exhibits a much more complicated time series pattern than the total mortality rate. The unemployment rate was remarkably low from 1950 to 1970. Within the seventies a dramatic

increase began that peaked in 1985 and 1986. After 1986 the unemployment rate began to decline up to 1990, then the trend reversed again. It seems that the oil crisis shocks that happened in the seventies initiated a restructuring process in the British economy and that the labour market changed substantially. From a time series perspective we have to acknowledge that the unemployment rate has a structural break, we suggest setting the breakpoint at the year 1970. The autocorrelation function calculated from the level values of the unemployment rate shows the same typical nonstationary, slowly decaying, shape that we have seen for the total mortality rate. When we look at the plot for the differenced unemployment rate we see the same kind of irregular but cyclical pattern: the amplitude of the cyclical pattern increases with time, starting in the seventies. The associated autocorrelation function for the first differences exhibits a more pronounced cyclical pattern, the maximum and borderline significant autocorrelation function value can be observed at lag 10. This period corresponds to a typical business cycle period. In general, time series with a pattern like the unemployment rate are difficult to deal with: the small number of observations will make it impossible to model the level time series separately for the two different subperiods. The first differences would require at least some kind of variance-stabilizing transformation, e.g., a Box-Cox transformation. For the exploratory analysis here, we take the level unemployment rate as it is and assume that the first differences are stationary.

3. **Spirits consumption per capita** (ALC), Fig. 4. The shape of the spirits consumption series is surprisingly very close to the shape of the level unemployment rate, although there is a slightly upward trend from the beginning in 1950. A dramatic increase of spirits consumption starts in 1970, the time series peaks in 1979. From 1980 to 1994 the series fluctuates around a very high level. The autocorrelation function is nearly identical to that of the unemployment rate. Looking at spirits consumption per capita in first differences, we see an abrupt increase in volatility from 1970 to 1980. This kind of heteroscedasticity is extremely difficult to handle. The autocorrelation function of the first differences looks like the autocorrelation function of a white noise process. In contrast to the first differences of the unemployment series no sign of a business cycle can be found.

4. **Cigarette consumption per capita** (CIG), Fig. 5. The cigarette consumption series also shows a structural break, but in contrast to the unemployment rate and to the spirits consumption series, cigarette consumption starts to decline dramatically

after 1970. From 1985 on, cigarette consumption starts to increase again. The autocorrelation function for the level series is again typical for a nonstationary time series. The first differences of the cigarette consumption series has no volatility outbreak but shows an increasing variance with time. The associated autocorrelation function also does not indicate any sign of a business cycle.

5. **Real Gross Domestic Product** (RGDP), Fig. 6. This variable is used as a simple socio-economic welfare indicator. At first glance, the real gross domestic product series shows a relatively smooth upward trend, but the plot of the series with first differences reveals a cyclical pattern, whereby the amplitude of the cycles increases dramatically from the seventies on. This cyclical pattern can also be identified in the autocorrelation function of the differenced time series, although it is not as strongly expressed as for the first differences of the unemployment series.

The description of the time series can now be summarized as follows:
Three of the time series, i.e., unemployment rate, spirits and cigarette consumption exhibit a structural break within the level series, the break-point year can be estimated roughly for all the three series at about 1970. With the exception of total mortality, all the time series exhibit heteroscedasticity in first differences. Two of the time series, i.e. unemployment rate and real gross domestic product, show a business cycle component. For the exploratory analysis here, we ignore the problems with structural breaks and heteroskedasticity and assume in addition that all time series are integrated of order one with stationary first differences.
The use of an age- and sex-adjusted mortality rate as dependent variable with non-adjusted independent variables is acceptable for our preliminary, exploratory analysis. The statistical problems of such an approach should be investigated thoroughly before any final conclusions can be drawn (Rosenbaum/Rubin [1984]).

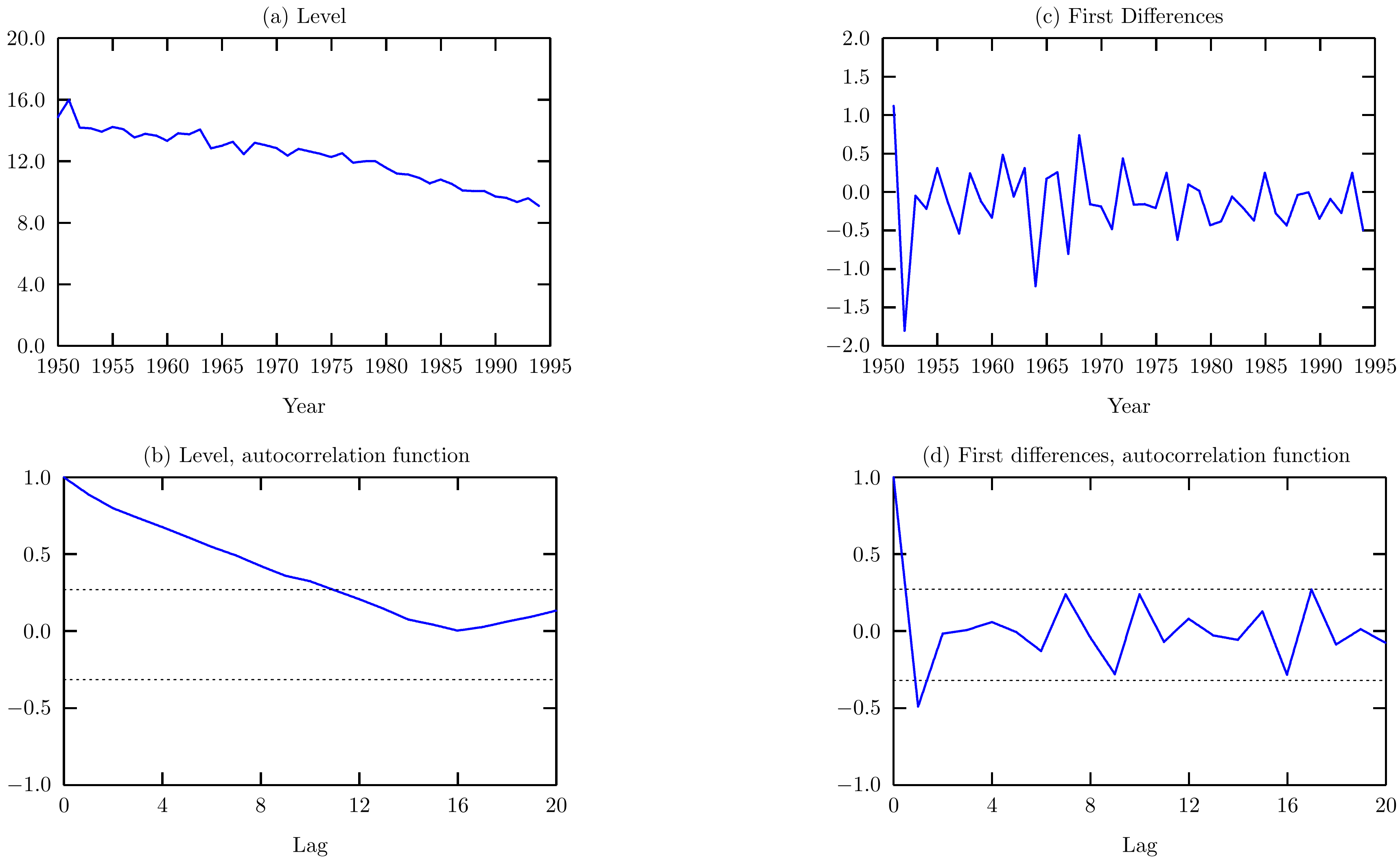


Figure 2: United Kingdom, Total Mortality Rate, per 1000: time series properties.

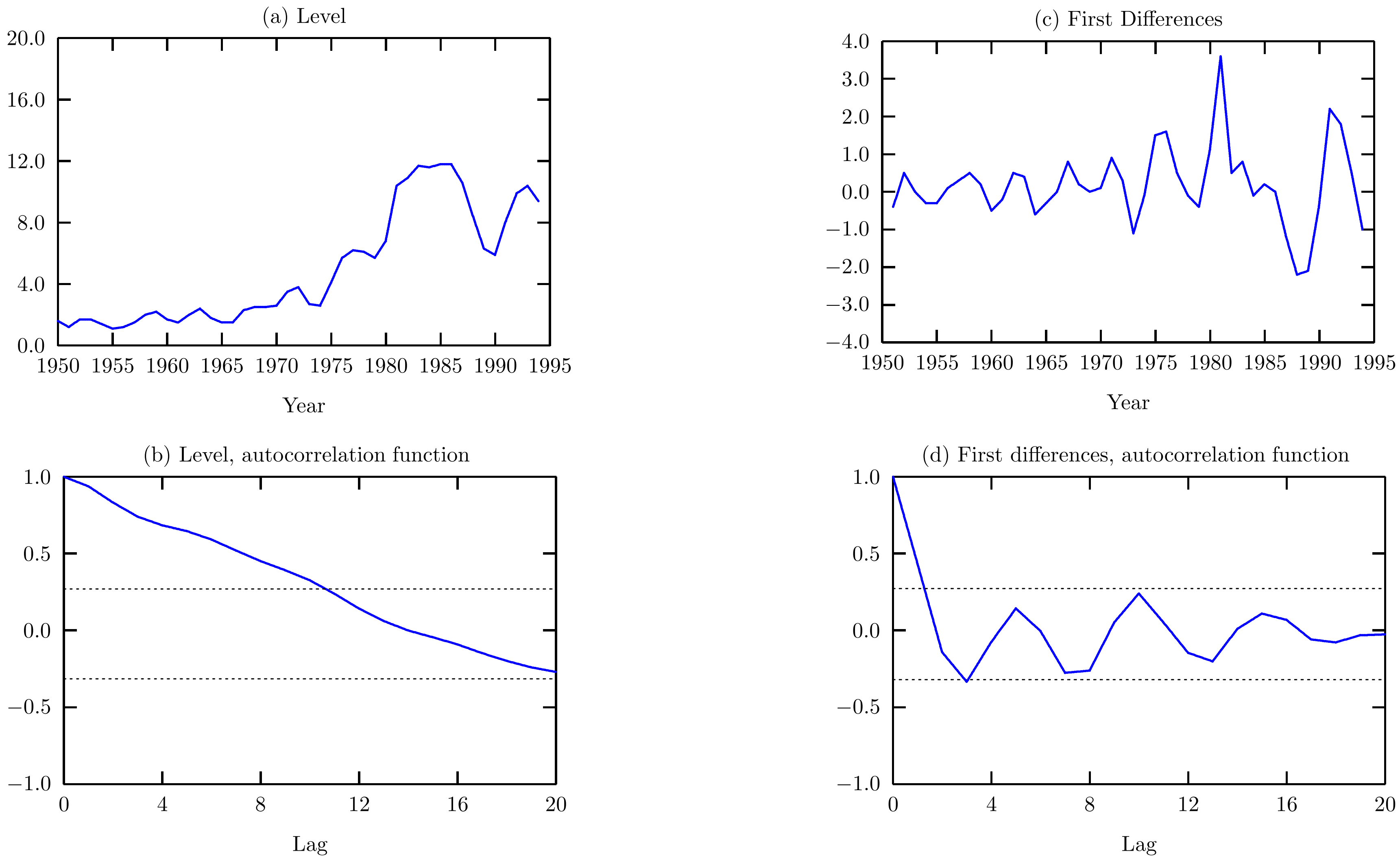


Figure 3: United Kingdom, Unemployment Rate, per 100: time series properties.

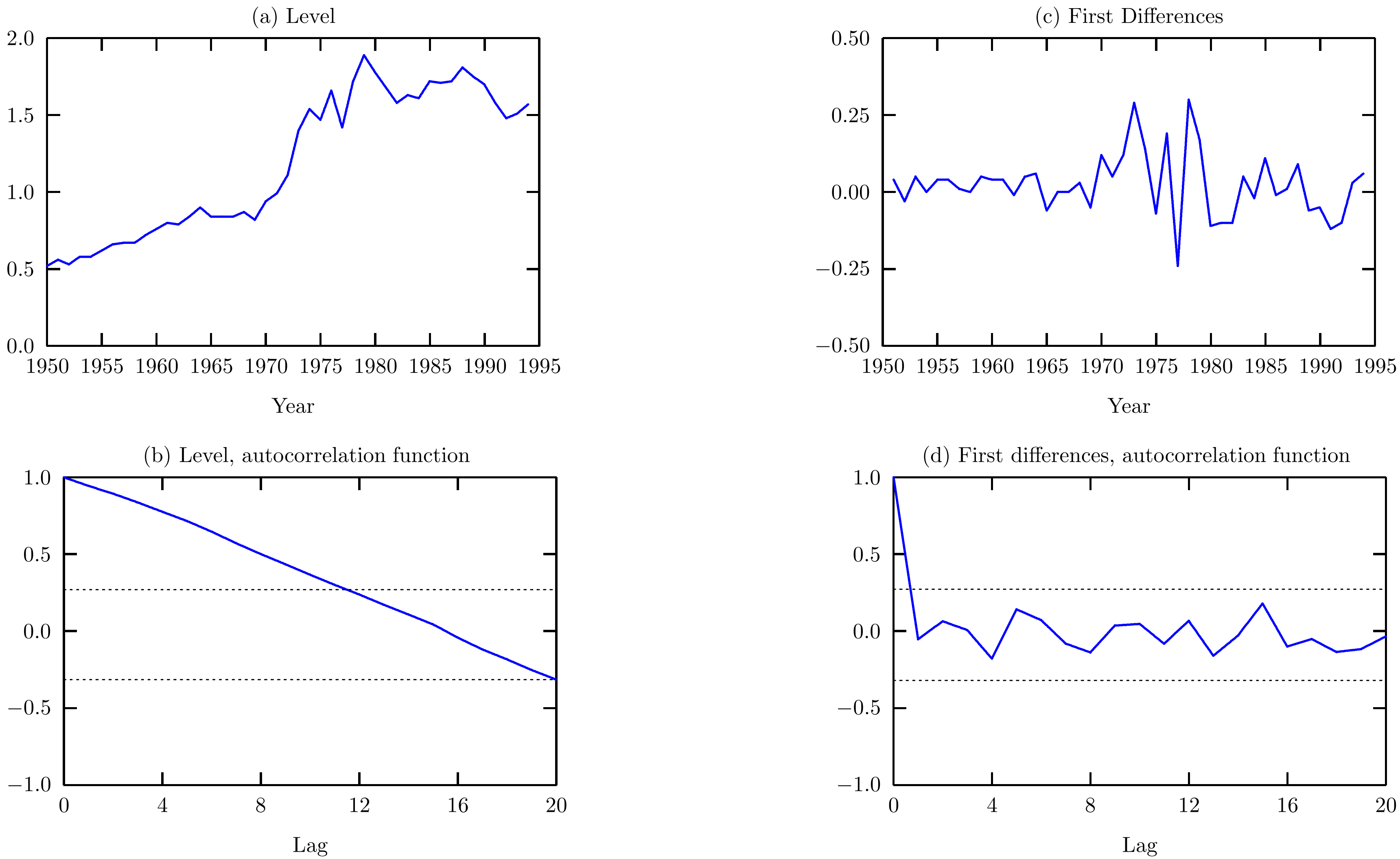


Figure 4: United Kingdom, Spirits consumption, liter per capita: time series properties.

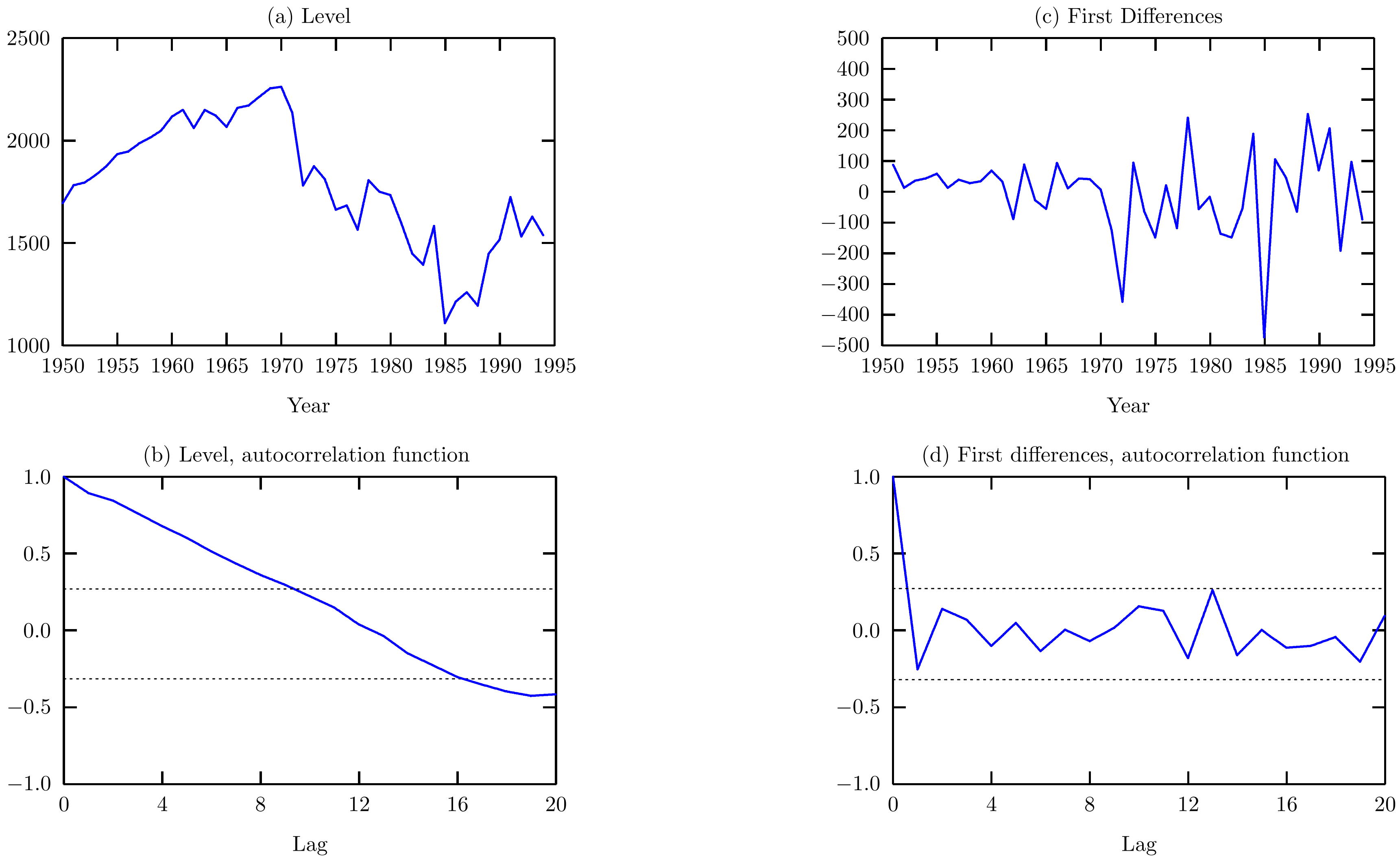


Figure 5: United Kingdom, Cigarettes per capita: time series properties.

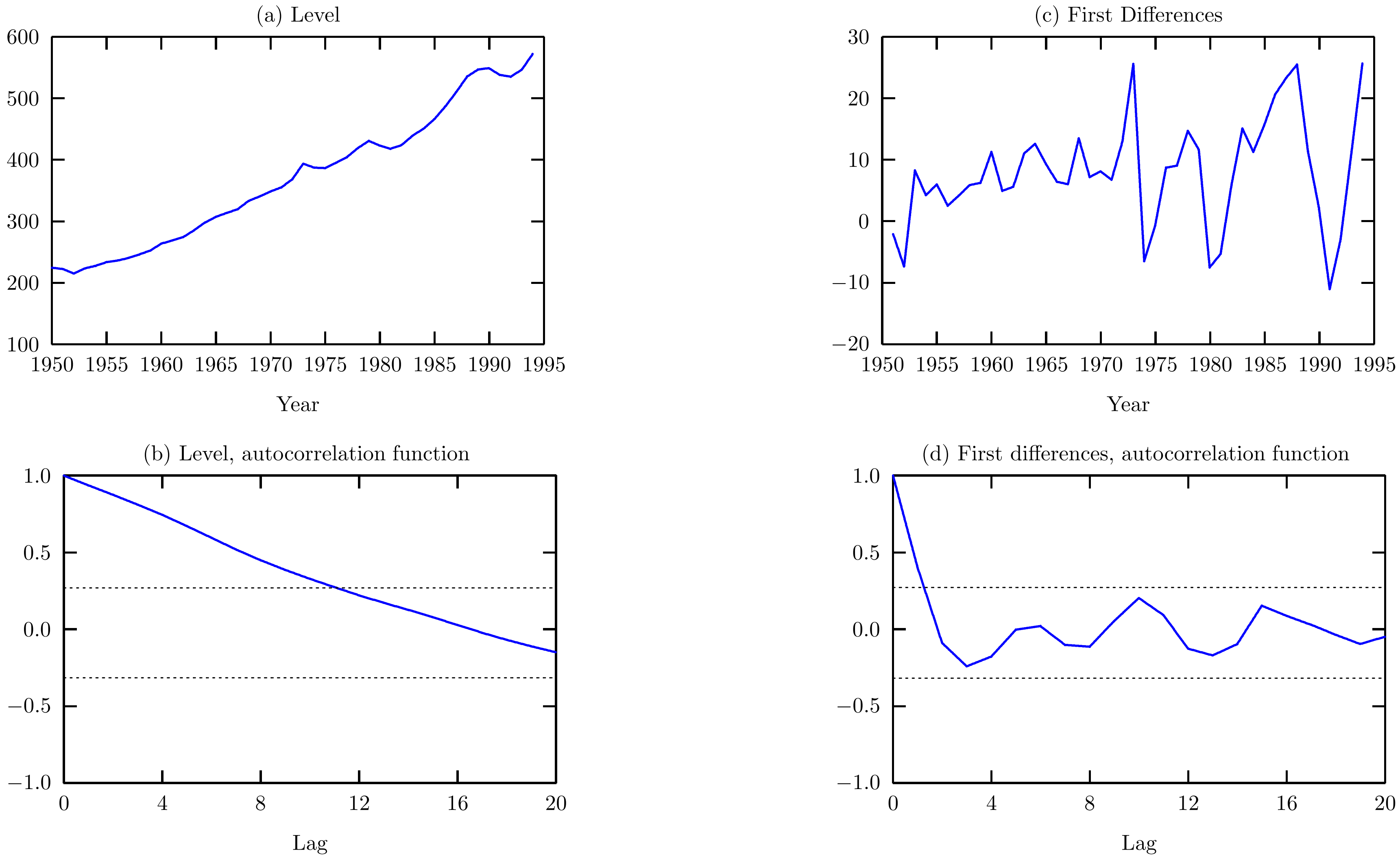


Figure 6: United Kingdom, Real Gross Domestic Product, 10^9 British Pounds, GDP deflator: 1990=100: time series properties.

2.2 Total mortality: estimated models

Based on the Poisson distributed lag approach we estimated 11 different models to analyze the relationship between total mortality and unemployment for the United Kingdom. In addition to the usual regression output we provide a plot of the estimated normalized lag distribution coefficients and a plot of the residuals autocorrelation function. The maximum lag length was set to 15 years.

Please note that when different Poisson lag distributions are to be compared, not only the plot of the normalized lag weights $w_{i,k}$ and the associated mean lag $\bar{\lambda}_k$ are relevant. The evaluation of the total effect also requires taking into account the associated parameter β_k, which is especially important because of the sign of β_k.

Models 1-4 are level models where the total mortality (RTMORT) is regressed on each of 4 independent variables, i.e. unemployment rate (UR), spirits consumption per capita (ALC), cigarette consumption per capita (CIG) and real gross domestic product (RGDP). The letter S is put at the beginning of a variable name when it refers to the sum of the weighted lagged variable, e.g.

$$\mathrm{SUR}_t(\lambda) = \sum_{i=0}^{15} w_i(\lambda)\mathrm{UR}_{t-i}.$$

All single variable models provide an excellent fit with a minimum R^2 of 0.924. The coefficients of the independent variable in each model are highly significant with p-values of less than 0.01. We are surprised to find negative signs not only for the coefficient of the unemployment rate variable but also for the spirits consumption variable. This would contradict our earlier findings. For real gross domestic product we find that the Poisson lag distribution is extremely right truncated and the mean lag $\bar{\lambda}$ is equal to 12.26 years, which would show a very long persisting impact of real gross domestic product on mortality. Right truncation also can be seen for the impact of spirits consumption but the truncation effect it is not as severe as for real gross domestic product. Despite the truncation we cannot increase the maximum lag length because with such a limited number of observations no direct comparison could be made with the multivariate models, where all 4 independent variables are included. The loss of degree of freedoms would no longer allow for valid statistical inference from such regressions.

Models 5-8 are used to repeat the first four regressions with all the variables in first differences. This is done as a check for spurious regression, which always may be a problem in time series regression analysis. The fit of all models drops dramatically downward, the R^2´s are close to zero and with the exception of spirits consumption the coefficients of the independent variable are no longer significant. These results indicate clearly that, for the derivation of useful results, all our time series regression analysis should be done within the framework and with the econometric tools of modern cointegration analysis, i.e. unit root tests and cointegration tests (e.g. Banerjee et al. [1993]). The discussion of these techniques is not within the scope of this introduction to distributed lag modeling. The main problem of applying these modern econometric tools is that the statistical characteristics of these procedures are not well defined when the data have anomalies like structural breaks and heteroskedasticity. But within the scope of our preliminary exploratory analysis, we can justify presenting the results of a so called 'error correction model' (Model 11), without being overly concerned by the statistical details.

Models 9 and 10 are multiple regression models where all 4 independent variables are included at the same time and all the parameters and lag distributions are estimated simultaneously. Model 9 uses all the variables in level, Model 10 uses all variables in first differences. In both cases the results are quite different from the results that we have found with the single independent variable models. In both models all the coefficients have the sign that we expected theoretically and confirmed the results of our earlier research work. When we compare the 'crude' and the 'adjusted' mean lags, we see that the most important changes occured when we compare the level variable models 1-4 and model 9: For the unemployment rate not only did the sign of the coefficient change but also the mean lag increased from 6.40 years to 9.48 years. The impact of spirits consumption changed the sign, and the mean lag was reduced from more than ten years to zero, i.e., there is only a contemporaneous effect of spirits consumption. The mean lag for cigarette consumption increased from 5.80 years to circa 10 years. The adjusted mean lag seems more plausible for cigarette consumption. In addition we see that the 'adjusted' lag distribution for cigarette consumption is also right truncated, which signals that the true lag distribution, which could be estimated when it would be possible to increase the maximum lag length, is even higher. For the real gross domestic product we observe that the 'adjusted' mean lag is much smaller than the 'crude' mean lag: it drops from more than 12 years to a little more than 2 years.

The lag distributions that could be derived from Model 10, where first differences had been used are very close to the lag distributions that we had found with the level variables. But in model 10 the fit, with an R^2 of 0.271, is not very good and the coefficients are only borderline significant with p-values smaller than 0.10.

Model 11 finally provides the results of an 'error correction model', which is essentially a first difference model where the discrepancy between the total mortality rate and the explained total mortality that was found in the level model 9 is added as an independent variable. The coefficient of this variable measures the speed of adjustment to 'equilibrium'. The use of an 'error correction model' makes the very strong assumption of a long-term stable equilibrium relationship between the levels of the model variables. The theoretical foundations for such a 'equilibrium' relationship are a subject of continueing discussions. We can see that all the lag distributions from model 9 to model 11 are very similar which gives us confidence to believe the overall results from an exploratory point of view. The short-term impact of adjusted spirits consumption raises the hypothesis that the main effect of spirits consumption on mortality in United Kingdom results in additional fatalities caused by accidents and violence. This finding should be confirmed by analysing cause-specific mortality rates.

The main result is a methodological one: we have to acknowledge that multiple regression models are necessary for the correction of potential confounders. Another important step of our future research will be the embedding of distributed lag analysis with several independent variables in the framework of cointegration analysis. That naive level time series regression models may result in spurious regressions is now known for more then 25 years (Granger/Newbold [1974]). Since further analysis of this subtle problem was given, e.g. by Philips[1986] and Plosser/Schwert[1986], the daily econometric routine of time series regression has changed dramatically. But even up to date, comprehensive econometric books (Banerjee et al. [1993]) provide no gold standard modeling approach. This is especially true when the data show additonal difficulties like stuctural breaks or heteroscedasticity. Both effects can be seen by our time series data from United Kingdom. In addition to these more standard problems, we cannot forget the unsolved epidemilogical time series difficulties which may arise from the use of age- und sex-adjusted variables in conjunction with non-adjusted variables.

Nevertheless, we are now focusing on a more narrow subproblem and summarize the main advantages of using Poisson distributed lags in comparison to Almon lags or Shiller lags as follows:

- Poisson distributed lags are characterized by one single parameter λ, that has a simple and natural interpretation as it is identical to the mean lag (as long as the lag distribution is not truncated).

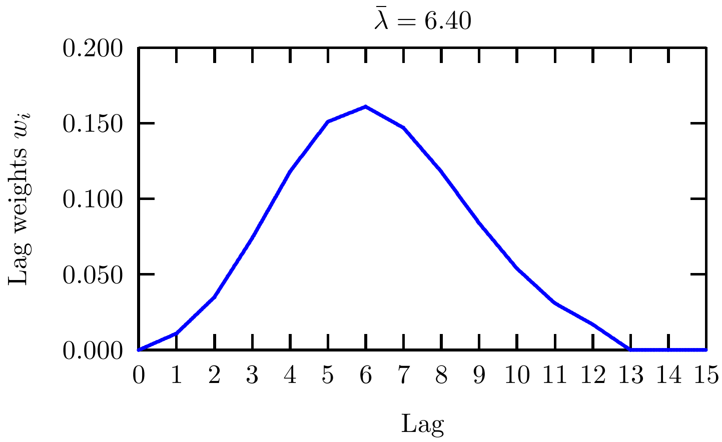
- Although the Poisson lags are determined by only one parameter λ, the functional form of Poisson lags is very flexible and the lag distributions are smooth functions, which enhances graphic presentations and comparisons.
- The comparison of different Poisson lag distributions is simple, and especially the epidemiologically important comparison of 'crude' versus 'adjusted' lag distributions is straightforward.

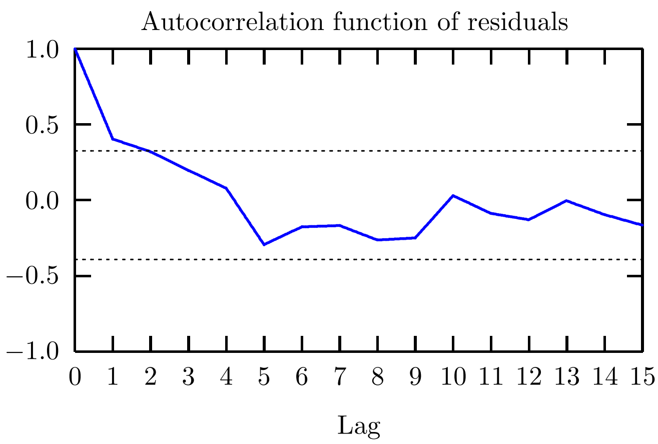
The main disadvantage of using Poisson lag distribution models is that the involved nonlinear optimization procedures may require substantial computation time.

Model 1: $\mathrm{RTMORT}_t = b_0 + b_1 \mathrm{SUR}_t(\lambda)$

Valid cases: 30		Dependent variable:	RTMORT
Missing cases:	0	Deletion method:	None
Total SS:	50.101	Degrees of freedom:	28
R-squared:	0.927	Rbar-squared:	0.924
Residual SS:	3.672	Std error of est:	0.362
F(1,28):	354.051	Probability of F:	0.000
Durbin-Watson:	0.984		

Variable	Estimate	Standard Error	t-value	Prob >|t|	Standardized Estimate	Cor with Dep Var
CONST	13.356046	0.121489	109.935900	0.000	---	---
SUR	-0.375132	0.019937	-18.816249	0.000	-0.962659	-0.962659



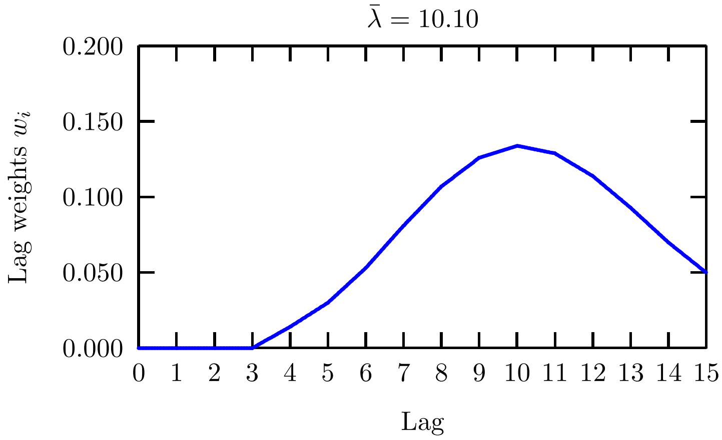


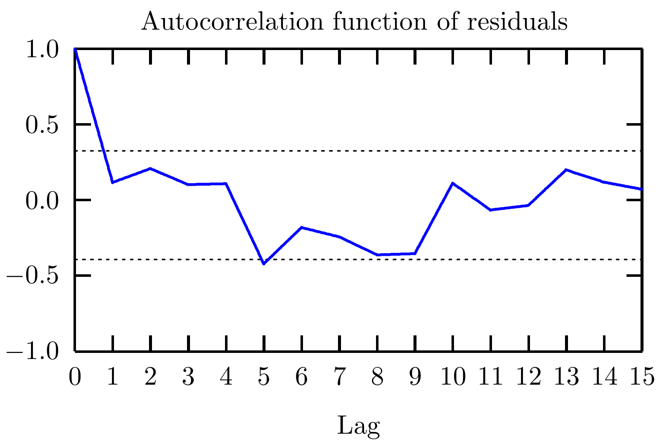
Modell 2: $\text{RTMORT}_t = b_0 + b_1 \text{SALC}_t(\lambda)$

```
Valid cases: 30                                   Dependent variable: RTMORT

Missing cases:                  0     Deletion method:                  None
Total SS:                  50.101     Degrees of freedom:                 28
R-squared:                  0.960     Rbar-squared:                    0.958
Residual SS:                2.009     Std error of est:                0.268
F(1,28):                  670.210     Probability of F:                0.000
Durbin-Watson:              1.626

                         Standard                 Prob   Standardized  Cor with
Variable     Estimate      Error      t-value     >|t|     Estimate     Dep Var
-------------------------------------------------------------------------------
CONST       15.104979   0.149843  100.805426     0.000       ---          ---
SALC        -3.241137   0.125196  -25.888406     0.000   -0.979744    -0.979744
```

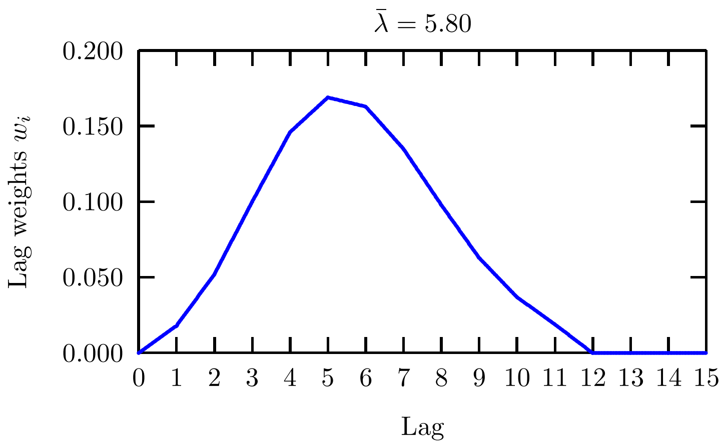


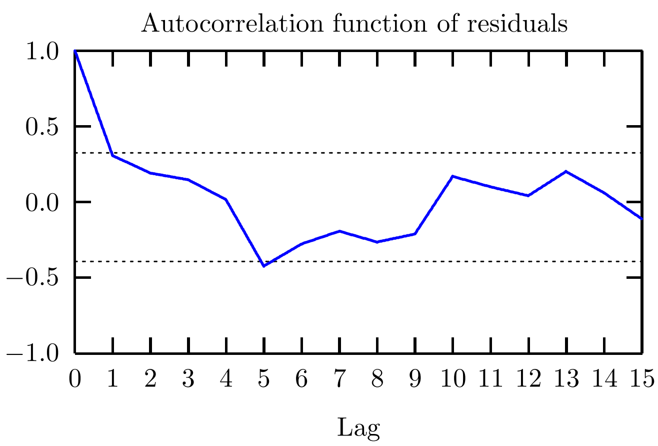


Model 3: $\text{RTMORT}_t = b_0 + b_1 \text{SCIG}_t(\lambda)$

Valid cases:	30	Dependent variable:	RTMORT
Missing cases:	0	Deletion method:	None
Total SS:	50.101	Degrees of freedom:	28
R-squared:	0.940	Rbar-squared:	0.938
Residual SS:	3.012	Std error of est:	0.328
F(1,28):	437.698	Probability of F:	0.000
Durbin-Watson:	1.179		

Variable	Estimate	Standard Error	t-value	Prob >|t|	Standardized Estimate	Cor with Dep Var
CONST	3.899955	0.365259	10.677231	0.000	---	---
SCIG	0.004159	0.000199	20.921244	0.000	0.969472	0.969472

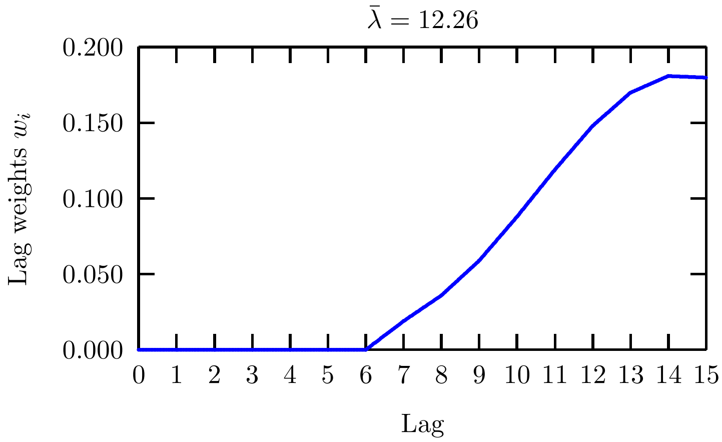


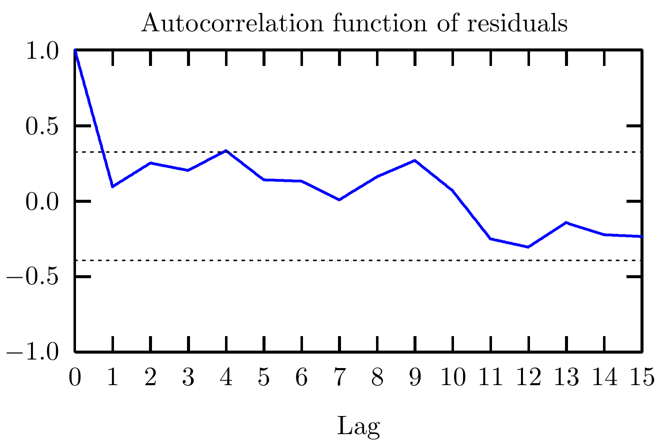


Model 4: $\text{RTMORT}_t = b_0 + b_1 \text{SRGDP}_t(\lambda)$

Valid cases:	30	Dependent variable:	RTMORT
Missing cases:	0	Deletion method:	None
Total SS:	50.101	Degrees of freedom:	28
R-squared:	0.964	Rbar-squared:	0.962
Residual SS:	1.822	Std error of est:	0.255
F(1,28):	741.941	Probability of F:	0.000
Durbin-Watson:	1.707		

Variable	Estimate	Standard Error	t-value	Prob >|t|	Standardized Estimate	Cor with Dep Var
KONST	17.455679	0.225772	77.315428	0.000	---	---
SRGDP	-0.018484	0.000679	-27.238597	0.000	-0.981648	-0.981648

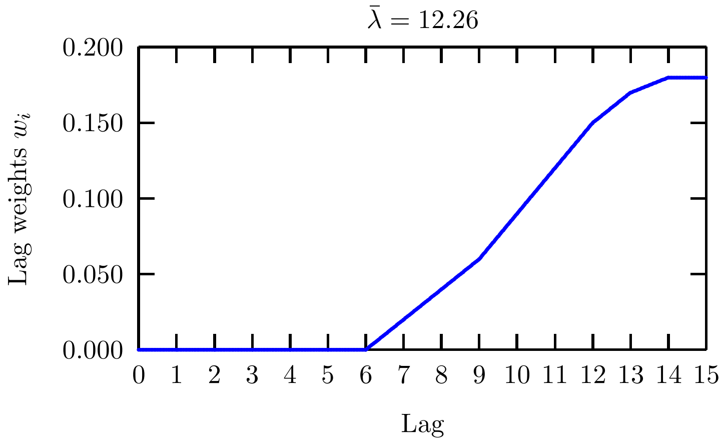


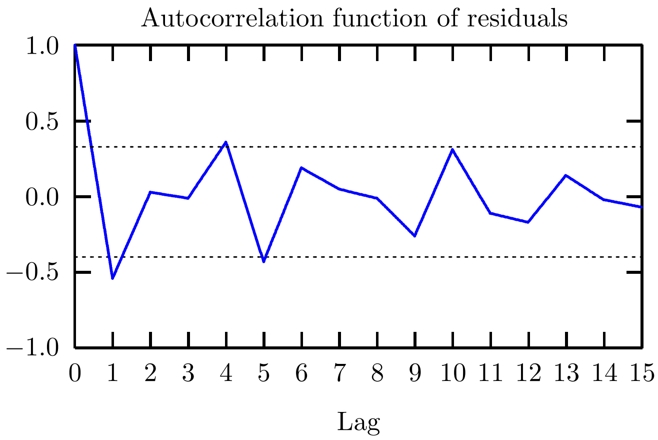


Model 5: $\text{DRTMORT}_t = b_0 + b_1\text{SDUR}_t(\lambda)$

Valid cases:	29	Dependent variable:	DRMORT
Missing cases:	0	Deletion method:	None
Total SS:	3.123	Degrees of freedom:	27
R-squared:	0.019	Rbar-squared:	-0.018
Residual SS:	3.065	Std error of est:	0.337
F(1,27):	0.511	Probability of F:	0.481
Durbin-Watson:	3.015		

Variable	Estimate	Standard Error	t-value	Prob >|t|	Standardized Estimate	Cor with Dep Var
CONST	-0.093901	0.084863	-1.106501	0.278	---	---
SDUR	-0.150368	0.210330	-0.714916	0.481	-0.136302	-0.136302

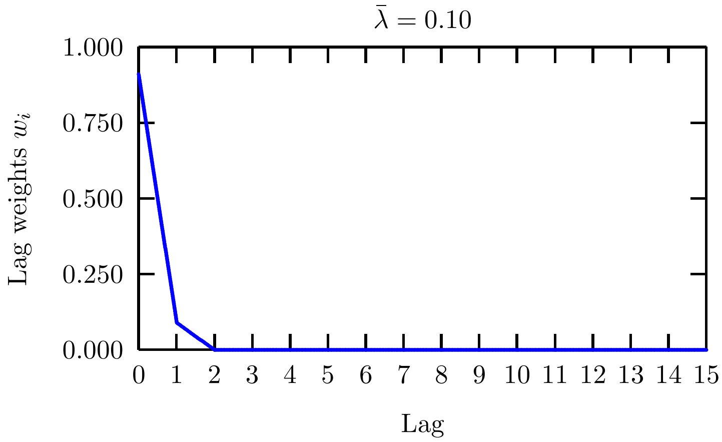


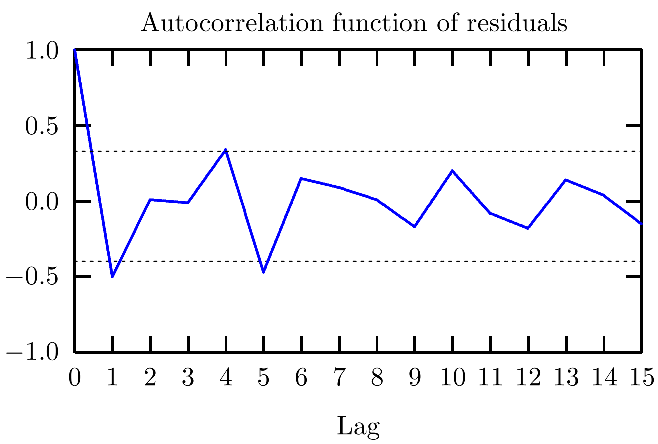


Model 6: $\text{DRTMORT}_t = b_0 + b_1\text{SDALC}_t(\lambda)$

Valid cases:	29	Dependent variable:	DRMORT
Missing cases:	0	Deletion method:	None
Total SS:	3.123	Degrees of freedom:	27
R-squared:	0.136	Rbar-squared:	0.104
Residual SS:	2.699	Std error of est:	0.316
F(1,27):	4.242	Probability of F:	0.049
Durbin-Watson:	2.867		

Variable	Estimate	Standard Error	t-value	Prob >|t|	Standardized Estimate	Cor with Dep Var
KONST	-0.162012	0.060172	-2.692474	0.012	---	---
SDALC	1.093878	0.531088	2.059693	0.049	0.368494	0.368494

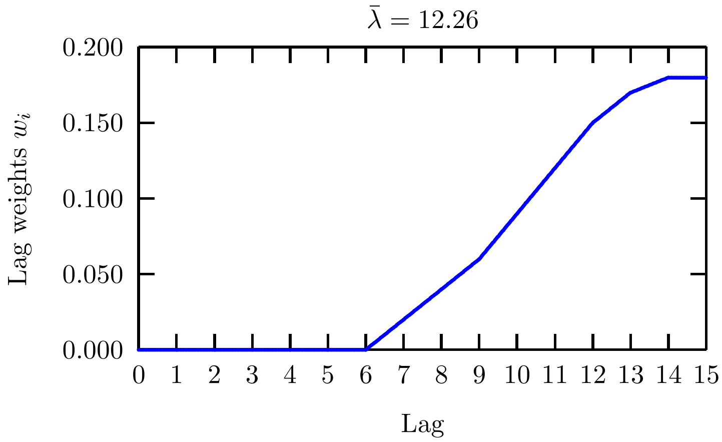


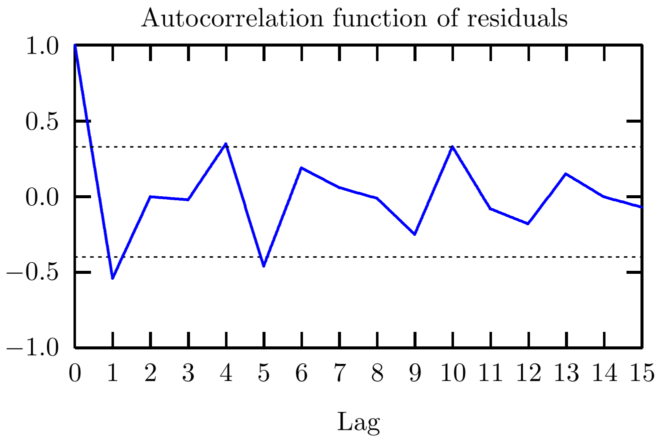


Model 7: $\text{DRTMORT}_t = b_0 + b_1\text{SDCIG}_t(\lambda)$

Valid cases:	29	Dependent variable:	DRMORT
Missing cases:	0	Deletion method:	None
Total SS:	3.123	Degrees of freedom:	27
R-squared:	0.039	Rbar-squared:	0.004
Residual SS:	3.001	Std error of est:	0.333
F(1,27):	1.104	Probability of F:	0.303
Durbin-Watson:	3.018		

Variable	Estimate	Standard Error	t-value	Prob >|t|	Standardized Estimate	Cor with Dep Var
KONST	-0.119803	0.063547	-1.885245	0.070	---	---
SDCIG	0.001625	0.001546	1.050782	0.303	0.198211	0.198211

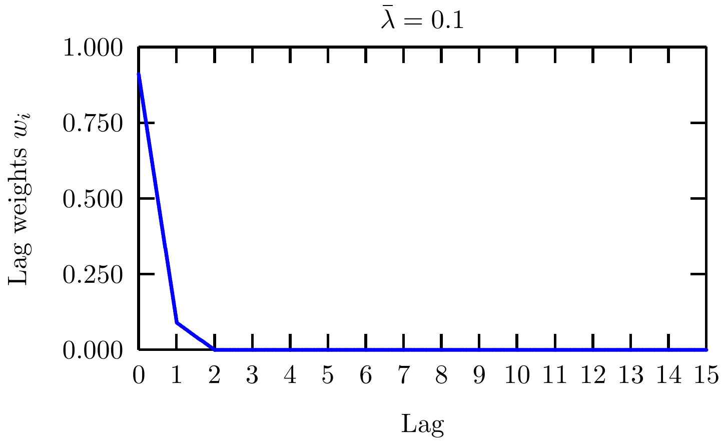


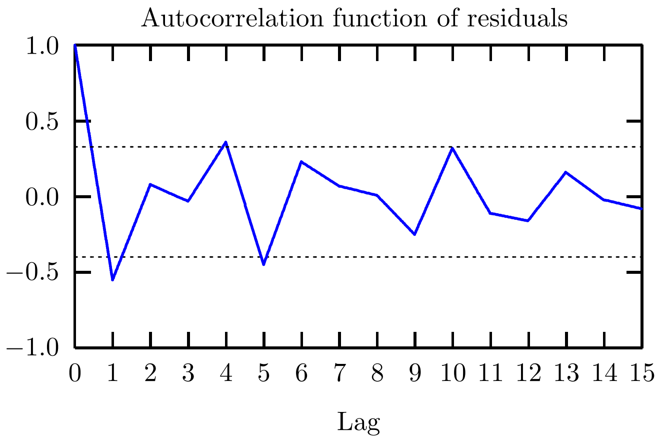


Model 8: $\text{DRTMORT}_t = b_0 + b_1\text{SDRGDP}_t(\lambda)$

Valid cases:	29	Dependent variable:	DRMORT
Missing cases:	0	Deletion method:	None
Total SS:	3.123	Degrees of freedom:	27
R-squared:	0.016	Rbar-squared:	-0.020
Residual SS:	3.072	Std error of est:	0.337
F(1,27):	0.450	Probability of F:	0.508
Durbin-Watson:	2.980		

Variable	Estimate	Standard Error	t-value	Prob >|t|	Standardized Estimate	Cor with Dep Var
KONST	-0.176098	0.087715	-2.007624	0.055	---	---
SDRGDP	0.004536	0.006759	0.671140	0.508	0.128097	0.128097

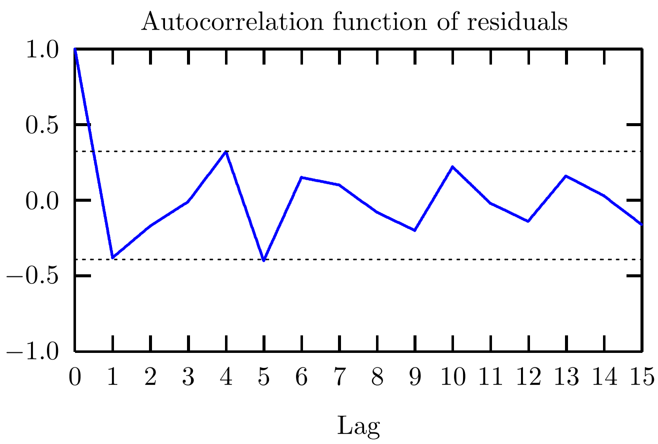




Model 9: $\text{RTMORT}_t = b_0 + b_1\text{SUR}_t(\lambda) + b_2\text{SALC}_t(\lambda) + b_3\text{SCIG}_t(\lambda) + b_4\text{SRGDP}_t(\lambda)$

Valid cases:	30	Dependent variable:	RTMORT
Missing cases:	0	Deletion method:	None
Total SS:	50.101	Degrees of freedom:	25
R-squared:	0.982	Rbar-squared:	0.980
Residual SS:	0.885	Std error of est:	0.188
F(4,25):	347.502	Probability of F:	0.000
Durbin-Watson:	2.740		

Variable	Estimate	Standard Error	t-value	Prob >|t|	Standardized Estimate	Cor with Dep Var
CONST	10.462199	1.146956	9.121713	0.000	---	---
SUR	0.710696	0.142506	4.987118	0.000	1.616201	-0.938987
SALC	0.823564	0.358819	2.295206	0.030	0.218419	-0.659281
SCIG	0.005171	0.000714	7.246578	0.000	0.913036	0.870176
SRGDP	-0.031662	0.004812	-6.579312	0.000	-1.909256	-0.968661



(Model 9 continued on next page)

(Modell 9 continued)

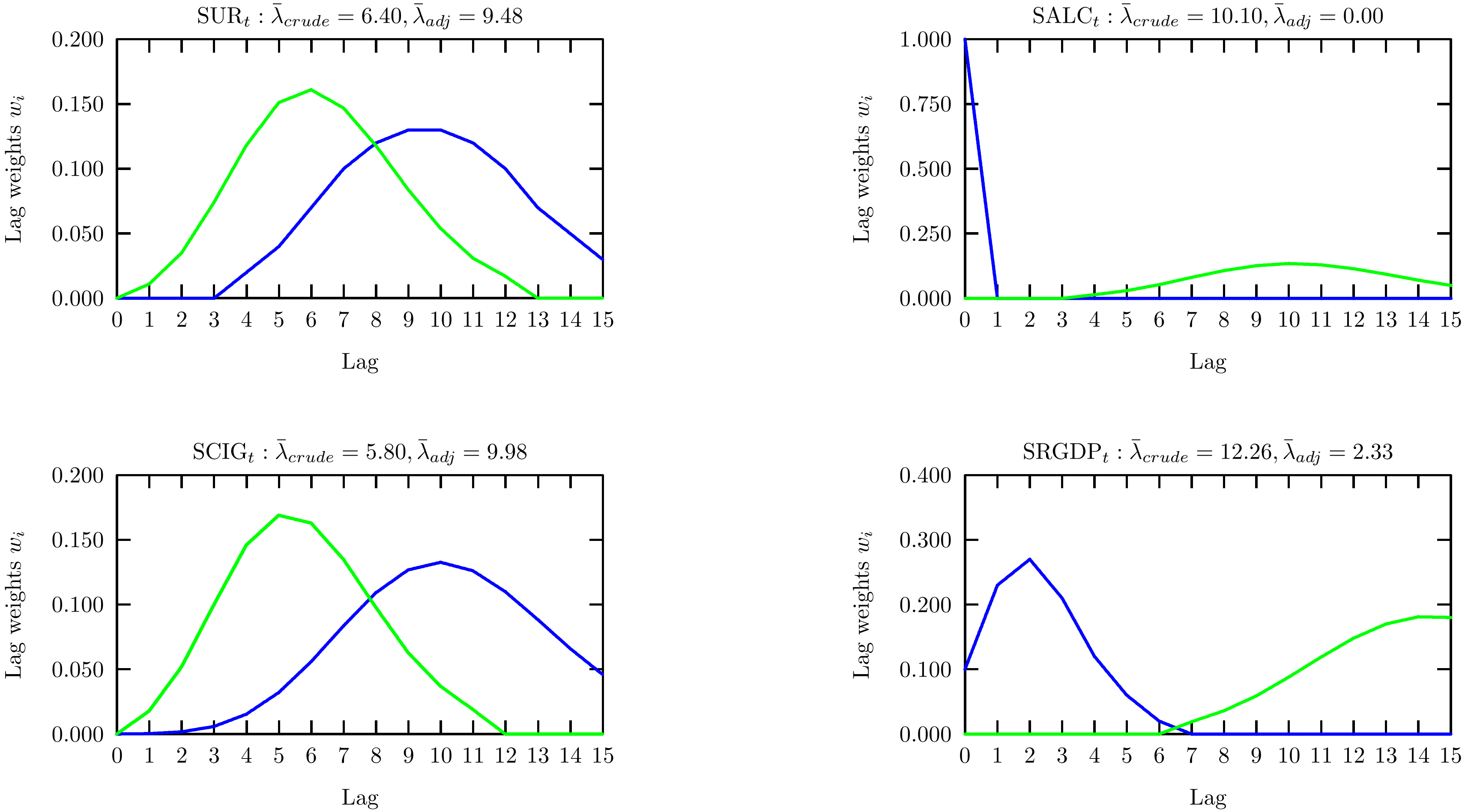


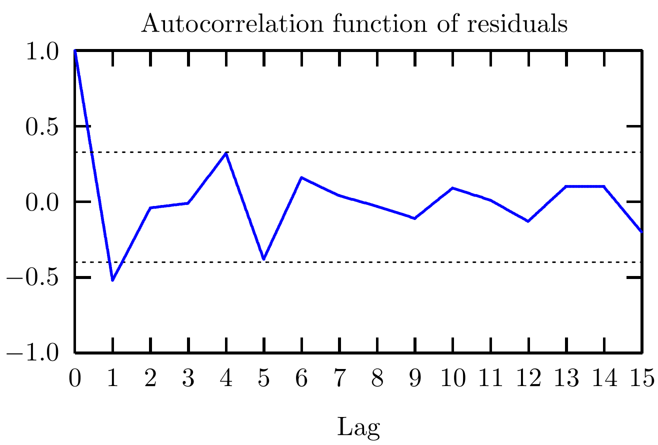
Figure 7: Level model, comparison of crude and adjusted Poisson lag distributions

Model 10:

$$\text{DRTMORT}_t = b_0 + b_1\text{SDUR}_t(\lambda_1) + b_2\text{SDALC}_t(\lambda_2) + b_3\text{SDCIG}_t(\lambda_3) + b_4\text{SDRGDP}_t(\lambda_4)$$

Valid cases:	29	Dependent variable:	RTMORT
Missing cases:	0	Deletion method:	None
Total SS:	3.123	Degrees of freedom:	24
R-squared:	0.271	Rbar-squared:	0.149
Residual SS:	2.277	Std error of est:	0.308
F(4,24):	2.229	Probability of F:	0.096
Durbin-Watson:	2.988		

Variable	Estimate	Standard Error	t-value	Prob >|t|	Standardized Estimate	Cor with Dep Var
KONST	0.012400	0.146242	0.084793	0.933	---	---
SDUR	0.952861	0.506232	1.882261	0.072	0.824401	-0.077014
SDALC	1.388260	0.532610	2.606525	0.015	0.513863	0.390671
SDCIG	0.006212	0.003285	1.891245	0.071	0.662503	0.190683
SDRGDP	-0.041622	0.023528	-1.769014	0.090	-0.606890	-0.012008



(Model 10 continued on next page)

(Modell 10 continued)

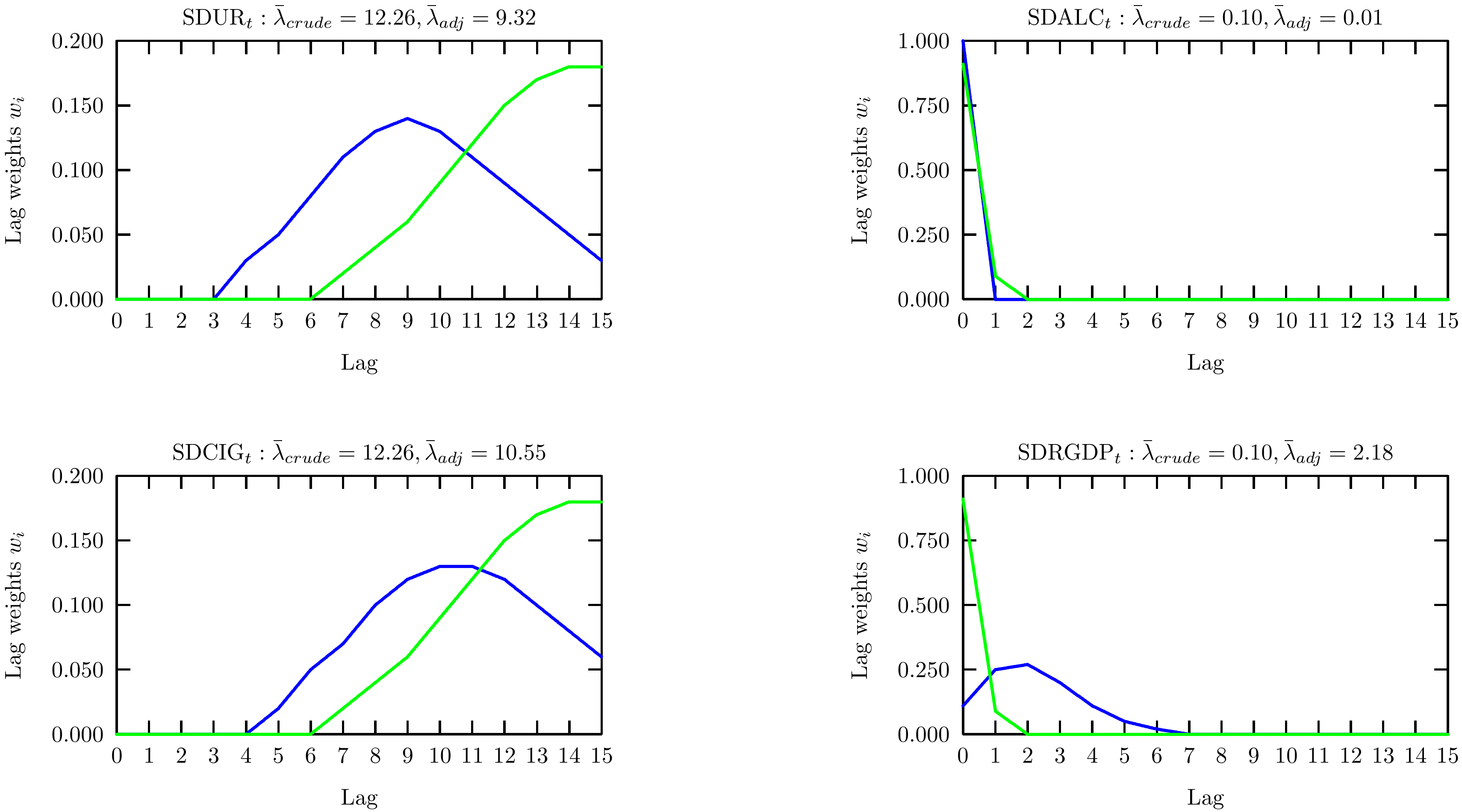


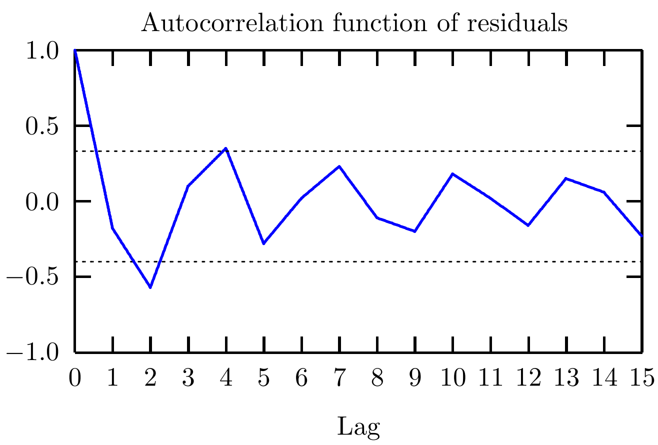
Figure 8: First difference model: comparison of crude and adjusted Poisson lag distributions

Model 11:

$$\mathrm{DRTMORT}_t = b_0 + b_1\mathrm{SDUR}_t(\lambda_1) + b_2\mathrm{SDALC}_t(\lambda_2) + b_3\mathrm{SDCIG}_t(\lambda_3) + b_4\mathrm{SDRGDP}_t(\lambda_4) + b_5\mathrm{ECM}_{t-1}$$

Valid cases:	29	Dependent variable:	DRTMORT
Missing cases:	0	Deletion method:	None
Total SS:	3.123	Degrees of freedom:	23
R-squared:	0.784	Rbar-squared:	0.737
Residual SS:	0.674	Std error of est:	0.171
F(5,23):	16.721	Probability of F:	0.000
Durbin-Watson:	2.318		

Variable	Estimate	Standard Error	t-value	Prob >|t|	Standardized Estimate	Cor with Dep Var
KONST	-0.014619	0.072424	-0.201858	0.842	---	---
SDUR	0.620757	0.242329	2.561626	0.017	0.537088	-0.077006
SDALC	0.915984	0.318145	2.879135	0.008	0.331703	0.386225
SDCIG	0.005141	0.001736	2.962339	0.007	0.509139	0.167537
SDRGDP	-0.025244	0.010141	-2.489370	0.020	-0.399378	-0.011911
ECM	-1.435687	0.186924	-7.680588	0.000	-0.762348	-0.796805



(Model 11 continued on next page)

(Modell 11 continued)

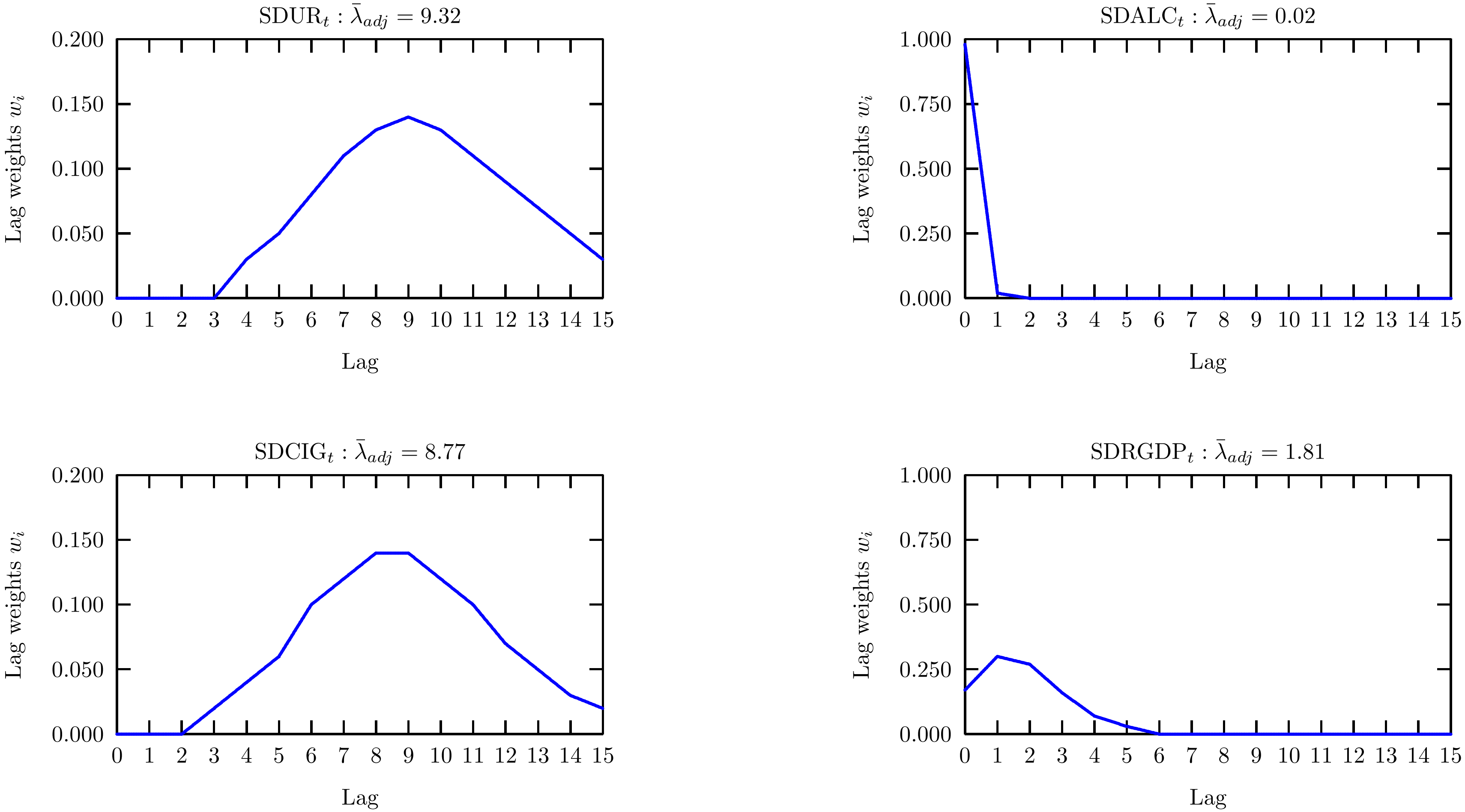


Figure 9: ECM Model, adjusted Poisson lag distributions

3 References

Akaike, H. [1974]: *A new look at statistical model identification.* IEEE Trans. Autom. Control, Vol. AC-19, S. 716-723

Almon, S. [1965]: *distributed lag between capital appropriations and expenditures.* Econometrica, Vol 33, pp. 178-196

Banerjee, A./Dolado J./Galbraith, J.W./Hendry D.F. [1993]: *Co-Intregration, Error-Correction, and the Econometric Analysis of Non-Stationary Data.* Oxford University Press

Dryhmes, P. J. [1971]: *Distributed Lags: Problems of Estimation and Formulation.* San Francisco, Holden Day

Fisher, I. [1937]: *Note on a short-cut method for calulating distributed lags.* Bulletin de l'Institut International de Statistique, Vol.29, pp. 323-327

Fahrion [1980]: *Endliche Lagstrukturen.* Wuerzburg

Friedrich, D. [1982]: *A Dynamic Adjustment Equation for a Poisson Distributed Lag Model.* Empirical Economics, Vol. 47, pp. 239-249

Goffe, W.L./Ferrier, G.D./Rogers, J. [1994]: *Global Optimization of Statistical Functions with Simulated Annealing.* Journal of Econometrics, Vol. 60, pp. 65-100

Grilliches, Z. [1967]: *Distributed Lags: A Survey.* Econometrica, Vol 35, pp. 16-49

Granger, C.W.J./Newbold, P. [1974]: *Spurious regressions in econometrics.* Journal of Econometrics, Vol. 2, 111-20

Judge, G.G./Griffits, W.E./Lütkepohl, H./Tsoung-Chao, L. [1985]: *Theory and Practice of Econometrics.* 2nd ed., John Wiley and Sons, New York

Murray, C.J.L./Chen, L.C. [1993]: *In search of a contemporary theory for understanding mortality change.* Soc. Sci. Med., Vol. 36, pp. 143-155

Phillips P.C.B. [1986]: *Understanding spurious regressions in econometrics.* Journal of Econometrics, Vol. 33, 311-340

Plosser C.I./Schwert G.W. [1986]: *Money, income and sunspots: Measuring economic relationships and the effects of differencing.* Journal of Monetary Economics, Vol. 4, 637-60

Nerlove, M. [1972]: *Lags in Economic Behavior.* Econometrica, Vol. 40, pp. 221-251

Schmidt, P. [1974]: *The small sample effect of various treatments of truncation remainders in the estimation of distributed lag models.* The Review of Economics and Statistics, Vol. 57, pp. 387-389

Schmidt, P./Guilkey, D. K. [1974]: *The effects of various treatments of truncation remainders in the estimation of distributed lag models.* Journal of Econometrics, Vol. 4, pp. 211-230

Rosenbaum, P.R./Rubin, D.B. [1984]: *Difficulties with Regression Analyses of Age-Adjusted Rates.* Biometrics, vol. 40, pp. 437-443

Shiller, R.J. [1973]: *A Distributed Lag Estimator Derived from Smoothness Priors.* Econometrica, Vol 41, pp. 775-788

Schwarz, G. [1978]: *Estimating the dimension of a model.* Ann. Stat., Vol. 6, S. 461-464

A European Reference Population

Age Group	Males	Females
0-4	16 175	15 352
5-9	16 542	15 715
10-14	17 292	16 460
15-19	18 417	17 508
20-24	19 859	18 974
25-29	19 963	19 265
30-34	19 001	18 514
35-39	18 120	17 821
40-44	17 195	16 991
45-49	14 590	14 636
50-54	14 444	14 839
55-59	13 430	14 246
60-64	12 200	13 937
65-69	10 069	13 256
70-74	5 957	8 653
75-79	5 419	9 019
80+	4 599	9 909

From:
United Nations: The Sex and Age Distribution of Population. The 1990 Revision of the United Nations Global Population Estimates and Projections. (Population Studies No. 122), New York 1991, p. 42, year 1990.

Das **Berliner Zentrum Public Health (BZPH)** ist eine Forschungs- und Serviceeinrichtung der drei Berliner Universitäten — Freie Universität Berlin, Humboldt-Universität zu Berlin, Technische Universität Berlin — in Kooperationmit außeruniversitäten Forschungseinrichtungen.

Das Berliner Zentrum Public Health nutzt die multidisziplinäre Vielfalt seiner Mitglieder, um komplexe Probleme praxisorientiert zu lösen.

Das Berliner Zentrum Public Health bietet Forschung, Beratung, Fortbildung und Serviceleistungen auf den Gebieten:

- Analyse des Gesundheitszustandes der Bevölkerung
- Stärkung der gesundheitlichen Kompetenz der Bevölkerung
- Entwicklung von Maßnahmen zur Gesundheitsförderung und Prävention
- umfassende Evaluation neuer Versorgungs- und Vernetzungsmodelle
- Effizienzsteigerung und Innovationen im Versorgungssystem
- Analyse und Weiterentwicklung der Strukturen im Gesundheitswesen.

Kontaktadresse: Berliner Zentrum Public Health
Geschäftsstelle
Ernst-Reuter-Platz 7
10587 Berlin

Authors:

Dr.-Ing. Günter Edenharter is Senior Econometrician in the Institute for Health Sciences at the Technical University of Berlin.

Prof. Dr. Dieter Friedrich is Professor of Statistics and Econometrics in the Institute for Business Informatics and Quantitative Methods at the Technical University of Berlin.

Prof. Dr. M. Harvey Brenner is Professor of Epidemiology in the Institute for Health Sciences at the Technical University of Berlin. He is also Professor of Health Policy and Management at the Johns Hopkins University, Baltimore, USA. Professor Brenner is Principal Investigator of the study for which this paper is one of the reports. The study is entitled: Estimating the social costs of unemployment and employment policies in the European Union and the United States, File No.: SOC-97-101881, for the European Commission Directorate General, Employment, Industrial Relations and Social Affairs V.G.2 - Employment Group.