DATA GRADUATION BASED ON STATISTICAL TIME SERIES METHODS.

Víctor M. Guerrero, Rodrigo Juárez and Pilar Poncela[*]

Abstract _____

Whittaker´s method is one of the most frequently employed techniques to graduate mortality tables. In order for the method to work and produce reasonable results, some subjective input is required from the graduator. In this paper we show that Whittaker´s solution to the graduation problem can be approached from a statistical time series model-based perspective that reduces the subjectivity in its application. It also serves to interpret the graduation problem as a classical estimation problem. In fact, on the basis of some suitable assumptions, we are able to show that the Best Linear Unbiased Estimator of the true mortality rates has the form of Whittaker´s solution. We also provide some complementary analytical tools aimed at helping the graduator to employ the method in practice and interpret its results from a statistical standpoint. A numerical illustration is shown in detail to exemplify the application of our proposal.

Keywords:
Best linear unbiased estimation; difference stationary processes; generalized least squares; robust techniques; Whittaker graduation.

[*]Guerrero, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Spain and Instituto Tecnológico Autónomo de México, 01000 México, D.F., MEXICO, e-mail: guerrero@gauss.rhon.itam.mx; Juárez, Instituto Tecnológico Autónomo de México; Poncela, Universidad Carlos III de Madrid, 126 28903 Getafe, Madrid.

Data graduation based on statistical time series methods

Víctor M. Guerrero[1,2], Rodrigo Juárez[2] and Pilar Poncela[1]

(1) Universidad Carlos III de Madrid, SPAIN and (2) Instituto Tecnológico Autónomo de México, MEXICO[1].

Abstract: Whittaker's method is one of the most frequently employed techniques to graduate mortality tables. In order for the method to work and produce reasonable results, some subjective input is required from the graduator. In this paper we show that Whittaker's solution to the graduation problem can be approached from a statistical time series model-based perspective that reduces the subjectivity in its application. It also serves to interpret the graduation problem as a classical estimation problem. In fact, on the basis of some suitable assumptions, we are able to show that the Best Linear Unbiased Estimator of the true mortality rates has the form of Whittaker's solution. We also provide some complementary analytical tools aimed at helping the graduator to employ the method in practice and interpret its results from a statistical standpoint. A numerical illustration is shown in detail to exemplify the application of our proposal.

Keywords: Best linear unbiased estimation; Difference stationary processes; Generalized least squares; Robust techniques; Whittaker graduation.

Abbreviated title: Graduation and time series.

---

[1]Correspondence to: Prof. Víctor M. Guerrero, Departamento de Estadística, Instituto Tecnológico Autónomo de México, 01000 México, D.F., MEXICO

1 Introduction

The problem of graduating a set of observed crude mortality rates has been approached from several standpoints (see London, 1985 for a general classification and description of methods). One of those approaches leads to the Whittaker graduation method, which basically consists of using a linear combination of the crude rates to obtain a series of graduated values. This new series should be close enough to the crude rates to satisfy a goodness of fit criterion and so smooth to be considered a better representation of the true ones. The trade off between goodness of fit and smoothness is controlled through a smoothing parameter that reflects the relative importance of those two aspects. Because of its intuitive appeal and relative ease of implementation, Whittaker graduation has been extensively employed in practice. However, both the standard to measure smoothness and the smoothing parameter itself are usually chosen arbitrarily, according to the graduator's experience. In order to get a better understanding of the method and appreciate the implications of those arbitrary decisions, Whittaker's method should be studied and interpreted within different settings.

Kimeldorf and Jones (1967) viewed Whittaker graduation from a Bayesian perspective and emphasized the use of prior distributions to incorporate the relevant knowledge that the graduator possesses before observing the data. In fact they advocated the use of a multivariate Normal distribution as a model for the vector of (correlated) true rates. Then they proceeded to suggest a way to postulate the covariance matrix from *a priori* arguments. Later on, Taylor (1992) elaborated on Kimeldorf and Jones' ideas and derived some results related to the smoothness standard and to the smoothing parameter involved. He considered a perspective in which the rates were assumed to be mutually independent. On that assumption, he interpreted the smoothing parameter as the reciprocal of a variance involved in his formulation. In Verrall (1993), the Bayesian interpretation was further exploited to pose the graduation problem in the form of a state space model. Such a formulation leads to the use of Kalman filtering to obtain the graduated values and allows estimation of the smoothing parameter from the observed data. However, the dynamic structure of the rates has to be decided by the analyst, which amounts to saying that the smoothness standard is chosen subjectively.

In this paper we suggest an alternative interpretation of Whittaker graduation that views the graduated series of rates as the Best Linear Unbiased Estimator (BLUE) of the true series. The smoothing parameter as well as the smoothness standard are then related to a dynamic specification derived from the observed data. Our focus will be placed on the statistical analysis of the data at hand. Therefore, some additional results are provided to enhance

the possibilities of analysis. In the following section we present the basic notation to be employed throughout this paper and briefly describe the basics of Whittaker graduation. Section 3 presents our theoretical proposal, which is supported by some assumptions there exposed. In Section 4 we show how to obtain a feasible solution and propose some further analyses that can be performed with the statistical tools presented in that section. In Section 5 we provide a numerical illustration aimed at getting some insight into how our proposed method works in practice. Then, Section 6 concludes the paper with some final remarks.

2   A brief description of Whittaker's graduation and basic notation

Let $\{u_x\}$ be a sequence of observed crude rates corresponding to ages $x = 1, \cdots, n$, where the actual age is given by $x + \alpha$, for some constant $\alpha > 0$. The basic idea underlying Whittaker graduation is to consider both fit and smoothness of the graduated rates through the following expression

$$M = \sum_{x=1}^{n} w_x(v_x - u_x)^2 + h \sum_{x=1}^{n-d} (\Delta^d v_x)^2 \qquad (2.1)$$

where v_x denotes a graduated value. Here, w_x is a non-negative weight, Δ is the forward differencing operator defined as $\Delta v_x = v_{x+1} - v_x$ for every v and x, d is the number of times the operator is applied (which implicitly defines the standard of smoothness) and h is a smoothing parameter that controls the relative importance of smoothness versus fit. The graduated values are those that minimize (2.1) for given d and h. Expression (2.1) can also be written in matrix notation as the quadratic form

$$M = (\mathbf{v} - \mathbf{u})'W(\mathbf{v} - \mathbf{u}) + h\mathbf{v}'S_d'S_d\mathbf{v} \qquad (2.2)$$

where $\mathbf{u} = (u_1, \cdots, u_n)$, $\mathbf{v} = (v_1, \cdots, v_n)$, $W = diag(w_1, \cdots, w_n)$ and S_d is an $(n - d) \times n$ matrix whose ij-th element is a binomial coefficient, i.e. it is of the form

$$S_d(i, j) = (-1)^{d+i-j} \begin{pmatrix} d \\ j - i \end{pmatrix} \quad \text{for } i = 1, \cdots, n - d \text{ and } j = 1, \cdots, n \quad (2.3)$$

with $\begin{pmatrix} d \\ k \end{pmatrix} = 0$ for $k < 0$ or $k > d$. The first summand in (2.2) measures fit and the second smoothness. The vector $\mathbf{v}$ that minimizes (2.2) can be shown

3

to be given by (see Shiu, 1986, for details)

$$\hat{\mathbf{v}} = (W + h S'_d S_d)^{-1} W \mathbf{u}. \tag{2.4}$$

This formula can be employed in practice once d and h are fixed. In fact, most practitioners would first fix the value d in 2, 3 or 4, according to their experience. Then h is selected by trial and error trading off smoothness against fit visually. It is thus evident that the practitioner's subjectivity plays an important role in choosing d and h.

Taylor (1992) relied on a Bayesian interpretation of Whittaker's method to propose a way for obtaining the value of h. His result was obtained on the assumption that the true values, $\{t_x\}$, are related to the observed $\{u_x\}$ through the expression

$$u_x = t_x + \epsilon_x \tag{2.5}$$

where ϵ_x is a zero-mean random error. $\{u_x\}$ was then assumed to be a sequence of independent and asymptotically Normally distributed random variables with $w_x = \text{var}^{-1}(u_x)$, and h was shown to be given by

$$h = \text{var}^{-1}(\Delta^d t_x). \tag{2.6}$$

However, in what follows we argue against the assumption of independence of the $\{u_x\}$.

We now realize that the value of d has implications on the behavior of the graduated rates. This can be seen in (2.2), since by minimizing $h \mathbf{v}' S'_d S_d \mathbf{v}$, the structure of $\{v_x\}$ tends to lie on an adaptive polynomial of degree $d - 1$. An implicit dynamic model is therefore employed when using Whittaker graduation. In the next section we propose using an explicit statistical formulation to take into account the potential autocorrelation structure of the rates. This is in clear opposition to the assumption of independence of $\{u_x\}$ mentioned above. We were motivated by some comments in Kimeldorf and Jones (1967) attributed to R. Henderson and M.D.W. Elphinstone, both of whom considered the rates to be connected, in the sense of having some kind of interrelationships. By assuming a specific dynamic structure among the rates we are able to select the value of d from the observed data and propose a new approach to graduation, based on well-known statistical (time series) methods. Our proposal is in line with Verrall's (1993) idea of employing a dynamic structure for the observed rates. However, we do not impose such a structure on the basis of *a priori* arguments, but rather use the observed data as a guide to select it.

The proposed time series representation is fairly general since it only assumes that, after appropriate differencing, a time series process becomes

stationary (see Box and Jenkins, 1976). That is,

$$\nabla^d T(Z_x) = a_x \qquad (2.7)$$

where ∇ is the backward differencing operator given by $\nabla Z_x = Z_x - Z_{x-1} = \Delta Z_{x-1}$, with $\nabla^d Z_x = \Delta^d Z_{x-d}$ for $d = 0, 1, \cdots$. Furthermore, $\{a_x\}$ is a zero-mean Gaussian stationary process. $T(\cdot)$ denotes an appropriate transformation, usually applied to make plausible the constant variance and Normal error (Gaussian) distribution assumptions.

3 Statistical model-based graduation

Our interest lies in estimating an unobserved series of rates $\{t_x\}$ that is known to follow a smooth pattern. A preliminary estimate of each (fix) parameter t_x is given by a crude rate u_x^0, obtained as the proportion of deaths out of n_x people exposed at age x. We start our derivations by considering the relationship (2.5) with u_x^0 in place of u_x, $\mathrm{E}(\epsilon_x) = 0$ and $\mathrm{var}(\epsilon_x) = t_x(1 - t_x)/n_x$, which come out from a binomial model with parameters t_x and n_x for each age x. Next, the fact that graduation is deemed neccessary may be due to some extraneous and undesirabe variability in the crude rates. The presence of outliers in this kind of data should then be considered reasonable. In what follows we shall allow for outliers by considering an outlier-corrected series $\{u_x\}$ obtained from $\{u_x^0\}$.

We now consider a log-odds-ratio transformation of the data, as did Verrall (1993), i.e.

$$T(u_x) = \log[u_x/(1 - u_x)] \ \text{ for } \ x = 1, \cdots, n \qquad (3.1)$$

which extends the range from $0 < u_x < 1$ to $-\infty < T(u_x) < \infty$. This is a monotone increasing transformation that will enable us to employ some well-known statistical methods. Its inverse is given by

$$T^{-1}(y) = \exp(y)/[1 + \exp(y)] \ \text{ for every } \ -\infty < y < \infty. \qquad (3.2)$$

Then, the dynamic structure in $\{u_x^0\}$ is assumed to be captured by the difference stationary representation (2.7). However, we consider that the potential autocorrelation structure in $\{\nabla^d T(u_x^0)\}$ is basically due to the presence of outliers.

3.1 Theoretical solution

We are now ready to assume the following.

Assumption 1. $\{T(u_x)\}$ admits the model

$$\nabla^d T(u_x) = a_{u,x} \text{ for } x = d + 1, \cdots, n \qquad (3.3)$$

with $T(u_1), \cdots, T(u_d)$ some fixed initial conditions and $\{a_{u,x}\}$ a zero-mean Gaussian white noise process.

As a consequence of using (3.1), the relationship (2.5) gets modified approximately as indicated by a first-order Taylor expansion. Thus we now assume.

Assumption 2. The transformed rates are related by means of

$$T(u_x) = T(t_x) + e_x \text{ for } x = 1, \cdots, n \qquad (3.4)$$

with $e_x \simeq [dT(u_x)/du_x]|_{u_x=t_x}(u_x - t_x)$, so that

$$\mathrm{E}(e_x) = 0 \text{ and } \mathrm{var}(e_x) = [n_x t_x (1 - t_x)]^{-1}. \qquad (3.5)$$

It should be clear that $\{T(u_x^0)\}$, $\{T(u_x)\}$ and $\{T(t_x)\}$ share the same long-run equilibrium, so that $\{e_x\}$ is stationary. In other words, the (transformed) series of rates are pairwise cointegrated, with cointegrating vector $(1, -1)$. Moreover, since the autocorrelation structure is supposedly due to outliers in the observed data, $\{e_x\}$ must follow a white noise process.

It is convenient to assume at this point that $\{T(t_x)\}$ follows a smooth pattern dictated by a representation that is in agreement with the cointegrating relationship aforementioned. This we do through the following.

Assumption 3. The dynamic structure of $\{T(t_x)\}$ gets captured by

$$\nabla^d T(t_x) = a_{t,x} \text{ for } x = d + 1, \cdots, n \qquad (3.6)$$

where $T(t_1), \cdots, T(t_d)$ are fixed initial conditions and $\{a_{t,x}\}$ is a zero-mean white noise process with $\mathrm{var}(a_{t,x}) = \sigma_t^2$ and uncorrelated with $\{e_x\}$

We remark that, on the one hand, Assumptions 1 and 2 are very natural and realistic given the nature of the data. On the other, Assumption 3 is a symplifying device that will lead us to a Whittaker-like solution. In fact, the latter assumption has been previously employed by other authors (e.g. Taylor, 1992) to incorporate a smoothness condition on $\{T(t_x)\}$.

We now define the column vectors $T(\mathbf{u}) = (T(u_1), \cdots, T(u_n))'$, $T(\mathbf{t}) = (T(t_1), \cdots, T(t_n))'$ and $\mathbf{e} = (e_1, \cdots, e_n)'$, where the prime indicates transpose, to express (3.4) as

$$T(\mathbf{u}) = T(\mathbf{t}) + \mathbf{e} \text{ with } \mathrm{E}(\mathbf{e}) = \mathbf{0} \text{ and } \mathrm{var}(\mathbf{e}) = W^{-1} \qquad (3.7)$$

where $W = diag(w_1, \cdots, w_n)$ and $w_x^{-1} = \mathrm{var}(e_x)$ for $x = 1, \cdots, n$. Similarly, by calling $\mathbf{a} = (a_{t,d+1}, \cdots, a_{t,n})'$, we can write (3.6) as

$$S_d T(\mathbf{t}) = \mathbf{a_t} \text{ with } \mathrm{E}(\mathbf{a_t}) = \mathbf{0}, \ \mathrm{var}(\mathbf{a_t}) = \sigma_t^2 I_{n-d} \text{ and } \mathrm{E}(\mathbf{a_t e'}) = 0. \qquad (3.8)$$

6

From (3.7) and (3.8) we construct the following system of equations

$$\left(\begin{array}{c} T(\mathbf{u}) \\ \mathbf{0} \end{array} \right) = \left(\begin{array}{c} I_n \\ S_d \end{array} \right) T(\mathbf{t}) + \left(\begin{array}{c} \mathbf{e} \\ -\mathbf{a}_t \end{array} \right) \tag{3.9}$$

where

$$E\left(\begin{array}{c} \mathbf{e} \\ -\mathbf{a}_t \end{array} \right) = \mathbf{0} \text{ and } \mathrm{var}\left(\begin{array}{c} \mathbf{e} \\ -\mathbf{a}_t \end{array} \right) = \left(\begin{array}{cc} W^{-1} & 0 \\ 0 & \sigma_t^2 I_{n-d} \end{array} \right) \tag{3.10}$$

Thus given $\mathbf{u}$, W, S_d and σ_t^2, we can apply Generalized Least Squares to obtain the BLUE of $T(\mathbf{t})$ as

$$\begin{aligned} \hat{T}(\mathbf{t}) &= \left[\left(\begin{array}{c} I_n \\ S_d \end{array} \right)' \left(\begin{array}{cc} W & 0 \\ 0 & \sigma_t^{-2} I_{n-d} \end{array} \right) \left(\begin{array}{c} I_n \\ S_d \end{array} \right) \right]^{-1} \\ &\times \left(\begin{array}{c} I_n \\ S_d \end{array} \right)' \left(\begin{array}{cc} W & 0 \\ 0 & \sigma_t^{-2} I_{n-d} \end{array} \right) \left(\begin{array}{c} T(\mathbf{u}) \\ 0 \end{array} \right) \\ &= (W + \sigma_t^{-2} S_d' S_d)^{-1} W T(\mathbf{u}) \end{aligned} \tag{3.11}$$

whose variance-covariance matrix $\Gamma = \mathrm{var}[(\hat{T}(\mathbf{t})]$ is given by

$$\begin{aligned} \Gamma &= (W + \sigma_t^{-2} S_d' S_d)^{-1} \\ &= W^{-1} - W^{-1}[W^{-1} + \sigma_t^{-2}(S_d^{0'} S_d^{0})^{-1}]^{-1} W^{-1}. \end{aligned} \tag{3.12}$$

We can easily see that (3.11) has the same form as Whittaker's solution (2.4) to the graduation problem in a transformed scale, when $h = \sigma_t^{-2}$ and the crude rates are corrected for outliers. We can also appreciate in (3.12) that Γ is an increasing function of σ_t^2, with $\Gamma \to 0$ as $\sigma_t^2 \to 0$ and $\Gamma \to W^{-1}$ as $\sigma_t^2 \to \infty$. Further, let us note that no distributional assumption was required to arrive at (3.11)-(3.12). If we now assume that $\{\mathbf{e}\}$ is Normally distributed, we get a Normal distribution for $\hat{T}(\mathbf{t})$, with mean $T(\mathbf{t})$ and variance-covariance matrix Γ. Hence, confidence intervals can also be constructed for $T(\mathbf{t})$.

3.2   Some complementary results

An aspect that can be exploited to get a better understanding of the role of σ_t^2 is that Γ^{-1} is the sum of two precisions, W and $\sigma_t^{-2} S_d' S_d$, associated to the binomial and the time series model, respectively. Then, as $\sigma_t^{-2} S_d' S_d$ corresponds to the smoothness element in (2.2), it is interesting to measure its precision share in Γ^{-1}. Guerrero (1993) proposed using to that end a scalar index derived by Theil (1963). A measure of the share of P in $(P + Q)^{-1}$, with P and Q positive definite matrices of size $n \times n$, is given by

$$\Lambda(P, P + Q) = \mathrm{tr}[P(P + Q)^{-1}]/n \tag{3.13}$$

Theil showed that this measure has the following properties:(i) it takes on values between zero and one, (ii) it is invariant under linear, nonsingular transformations of the variable involved, (iii) it behaves linearly, and (iv) $\Lambda(P, P+Q) + \Lambda(Q, P+Q) = 1$. The share of $\sigma_t^{-2} S_d' S_d$ in Γ^{-1} is thus measured as

$$
\begin{aligned}
\Lambda(\sigma_t^{-2} S_d' S_d, \Gamma^{-1}) &= \operatorname{tr}[\sigma_t^{-2} S_d' S_d (W + \sigma_t^{-2} S_d' S_d)^{-1}]/n \\
&= \operatorname{tr}\{[\sigma_t^{-2} W (S_d' S_d)^{-1} + I_n]^{-1}]\}/n
\end{aligned}
\tag{3.14}
$$

and we observe that it is a decreasing function of σ_t^2, with

$$
0 < \Lambda(\sigma_t^{-2} S_d' S_d, \Gamma^{-1}) \le 1
\tag{3.15}
$$

We call this an index of precision share attributable to the time series model since it can also be expressed as

$$
\Lambda(\sigma_t^{-2} S_d' S_d, \Gamma^{-1}) = \operatorname{tr}(\{\operatorname{var}^{-1}[T(\mathbf{u})] - \operatorname{var}^{-1}[\hat{T}(\mathbf{t})]\}\operatorname{var}[\hat{T}(\mathbf{t})])/n
\tag{3.16}
$$

so that it is associated to the relative reduction in variance attributable to using $\hat{T}(\mathbf{t})$ rather than just $T(\mathbf{u})$. Thus, given $\hat{\sigma}_t^2$ we can calculate the precision share achieved with such a choice. Moreover, we can decide the value of σ_t^2 by fixing a desirable precision share and then selecting $\hat{\sigma}_t^2$ as the value that satisfies such a condition. However, the empirical validation of the model assumptions should also be kept in mind and an extremely smooth estimated series may not be supported by the data at hand.

Next, let us recall that in the transformed scale (where the errors are Normally distributed) $T(t_x) = \operatorname{E}[\hat{T}(t_x)]$, so that $T(t_x)$ is also the median of $\hat{T}(t_x)$. Then if we apply the inverse transformation to bring the estimator to the original scale, we obtain

$$
\hat{t}_x = T^{-1}[\hat{T}(t_x)] \quad \text{for} \quad x = 1, \cdots, n
\tag{3.17}
$$

which (by monotonicity of $T(\cdot)$) has its median located at t_x. The whole series of estimates $\{\hat{t}_x\}$ is what we call the graduated rates with our proposal. Besides, confidence limits can also be constructed in the original scale by applying the inverse transformation to the limits obtained in the transformed scale.

4 An empirically supported working solution

In order to apply the previous results we should determine the values of S_d, W and σ_t^2 from the observed data in the most objective way. As a

starting point we need to correct the data for outliers, since some part of the oberved roughness is assumed to be due to them. If the data follow a Normal distribution, except for a small fraction coming form a heavy tailed symmetric distribution, we can use Huber's function, given for $z \sim N(0,1)$ by

$$\psi_c(z) = min\{c, max\{z, -c\}\} = \begin{cases} z & \text{if } |z| \leq c \\ c\, Sgn(z) & \text{if } |z| > c. \end{cases} \quad (4.1)$$

This function produces robust location estimates that are optimal in a minimax sense. In fact they are Maximun Likekihood estimates for the least favorable distribution (see Hampel et al., 1986, for more information on this topic). The constant c depends on the expected fraction of contaminated data and is usually chosen as $c = 1.645$, thus considering a potential 10% contamination.

In our case, Huber's function will be used for robust estimation of dynamic location. Since for each age x there is only one observation, it is impossible to judge its degree of contamination when considering the ages separately. Nevertheless, when $\{T(u_x^0)\}$ is viewed as a whole, its dynamic structure sets the standard for comparison. This structure is completely determined by the value of d that renders the series stationary. Hence, to specify d we can use an Augmented Dickey-Fuller (ADF) test on $\{T(u_x^0)\}$ (see Hamilton, 1994, for a detailed description of this test and related ones). Once d has been fixed, S_d can be constructed as indicated by (2.3). The elements of W can be preliminarily estimated by replacing t_x by u_x^0 in (3.5) to obtain the consistent estimates

$$\tilde{w}_x = \tilde{var}^{-1}(e_x) = n_x u_x^0(1 - u_x^0) \quad \text{for } x = 1, \cdots, n \quad (4.2)$$

so that $\tilde{W} = diag(\tilde{w}_1, \cdots, \tilde{w}_n)$. Then the value of $\hat{\sigma}_t^2$ can be decided from (3.14) by fixing a desirable (large) precision share attributable to the time series model and then looking for the corresponding σ_t^2 that produces it.

The last point that deserves special attention from a statistical analysis standpoint, refers to validating the model assumptions. In terms of the random errors, such assumptions are the following. (1) $e_x = T(u_x) - T(t_x) \sim N(0, w_x^{-1/2})$ and (2) $\{a_{t,x}\}$ is a zero-mean white noise process with variance σ_t^2 and uncorrelated with $\{e_x\}$, where $a_{t,x} = \nabla^d T(t_x)$. Then it is natural to calculate the residuals $\hat{e}_x = T(u_x) - \hat{T}(t_x)$ to verify (1). Similarly, we may be tempted to calculate $\hat{a}_{t,x} = \nabla^d \hat{T}(t_x)$, but let us notice that this is not very natural, since no observations would then be employed to obtain $\{\hat{a}_{t,x}\}$. In fact, we should realize that (3.11) implies

$$W[T(\mathbf{u}) - \hat{T}(\mathbf{t})] = \sigma_t^{-2} S_d' S_d \hat{T}(\mathbf{t}) \quad (4.3)$$

9

so that $\hat{\mathbf{e}} = T(\mathbf{u}) - \hat{T}(\mathbf{t})$ and $S_d\hat{T}(\mathbf{t}) = \mathbf{a}_t$ would lead to

$$W\hat{\mathbf{e}} = \sigma_t^{-2}S_d'\hat{\mathbf{a}}_t. \tag{4.4}$$

Therefore, if we have $\hat{\mathbf{e}} \sim N(\mathbf{0}, W^{-1})$ and $\hat{\mathbf{a}}_t \sim N(0, \sigma_t^2 I)$, as required by the model assumptions, (4.4) would imply that $W = \sigma_t^{-2}S_d'S_d$, which is not possible because W is diagonal with all its elements different from each other. Even if it were possible, then we would have $\hat{T}(\mathbf{t}) = \frac{1}{2}T(\mathbf{u}) = \hat{\mathbf{e}}$ and $S_d\hat{\mathbf{e}} = \hat{\mathbf{a}}_t$ so that $E(\hat{\mathbf{a}}_t\hat{\mathbf{e}}') = S_dW^{-1} \neq 0$. Our conclusion is that the sole definition of $\hat{\mathbf{a}}_t$ is not supported by the data. That is why we should consider Assumption 3, and (3.6) in particular, as a convenient and simplifying device that helps to obtain reasonable results, although it is not testable empirically. On the other hand, validating the assumptions on $\{e_x\}$ through a residual analysis does not pose any difficulty, as it will be appreciated in the numerical illustration. We should also be aware of the usefulness of a more flexible transformation than (3.1) to ensure that the assumptions are valid in data. Thus we suggest to keep in mind a Box-Cox power transformation of the odds-ratios, when needed, that is

$$T_\lambda(u_x) = [u_x/(1 - u_x)]^{(\lambda)} \tag{4.5}$$

where, for every variable Y, we have

$$Y^{(\lambda)} = \begin{cases} (Y^\lambda - 1)/\lambda & \text{if } \lambda \neq 0 \\ log(Y) & \text{if } \lambda = 0. \end{cases} \tag{4.6}$$

Hence, when $\lambda = 0$, (4.6) yields (3.1). See Guerrero and Johnson (1982) for more information about the use of (4.6) in some other contexts.

The following steps summarize our proposal:

Step 1 Determine the degree of differencing from an ADF test on $\{T(u_x^0)\}$ and create S_d. Obtain $\tilde{W} = diag(\tilde{w}_1, \cdots, \tilde{w}_n)$ with $\tilde{w}_x = n_x u_x^0(1 - u_x^0)$ for $x = 1, \cdots, n$.

Step 2 Fix the value $\hat{\sigma}_t^2$ in such a way that $\Lambda(\hat{\sigma}_t^{-2}S_d'S_d, \Gamma^{-1}) = \eta$ with $0 < \eta < 1$ specified beforehand, where $\Lambda(\cdot, \cdot)$ is given by (3.14).

Step 3 Calculate the preliminary values of $\tilde{T}(\mathbf{t})$ and $\tilde{\Gamma}$ with (3.11)-(3.12).

Step 4 Obtain the following outlier-corrected results, for $x = 1, .., , n$. (i) residuals $\tilde{e}_x = \tilde{w}_x^{-1/2}\psi_{1.645}\{\tilde{w}_x^{1/2}[T(u_x^0) - \tilde{T}(t_x)]\}$; (ii) observations $T(u_x) = \tilde{T}(t_x) + \tilde{e}_x$ and $u_x = T^{-1}[T(u_x)]$; and (iii) weights $\hat{w}_x = n_x u_x(1 - u_x)$.

Step 5 Calculate the final values of $\hat{T}(\mathbf{t})$, $\hat{\Gamma}$ and $\Lambda(\hat{\sigma}_t^{-2}S_d'S_d, \hat{\Gamma}^{-1})$.

10

Step 6 Carry out a residual analysis on $\{\hat{w}_x^{1/2}\hat{e}_x\}$. If the assumptions are supported by the data, you are finished, otherwise go back to step 2 and try with another value of η. As a guide, the higher the value of η, the less autocorrelation in the standardized residuals, but the more non-Normal (platikurtic) its distribution.

5   A numerical illustration

Here we exemplify the application of our proposal by graduating an actual series of crude mortality rates. The data set used for this example is the 1971 Group Annuity Mortality Table for Female Lives appearing in London (1985). As a preliminary step, we present in Figure 1 the original data and reproduce the results obtained by London using Whittaker graduation. In this figure, the numbers in parentheses (1000 and 426138) are two values of the smoothing parameter h, employed originally by London. The graduation was first carried out in the logarithmic scale assuming $d = 2$ and the graduated rates were then brought back to the original scale by taking anti-logs. Visual appreciation of the data (crude and graduate rates) is the key element to make decisions and judge the adequacy of these results.

FIGURE 1 AROUND HERE

To apply our suggested procedure we start by using the ADF test on the observed (transformed) rates. In this case, since there may be some doubt whether $d = 1$ or $d = 2$ is the appropriate degree of differencing, we postulate the model

$$\nabla T(u_x^0) = \rho \nabla T(u_{x-1}^0) + \beta_1 \nabla^2 T(u_{x-1}^0) + \cdots + \beta_{p-1} \nabla^2 T(u_{x-p+1}^0) + \epsilon_x$$

so that testing $H_0 : d = 2$ vs. $H_1 : d = 1$ is equivalent to testing $H_0 : \rho = 1$ vs. $H_1 : \rho < 1$. We chose initially $p = 10$ and the final value $p = 7$ was decided by discarding the highest lag in a regression when its coefficient was not significantly different from zero at the 10% level. The following results were obtained.

$$\rho = \begin{matrix} 0.923, \\ (0.457) \end{matrix} \quad \hat{\beta}_1 = \begin{matrix} -1.606, \\ (0.461) \end{matrix} \quad \hat{\beta}_2 = \begin{matrix} -1.798, \\ (0.463) \end{matrix} \quad \hat{\beta}_3 = \begin{matrix} -1.822, \\ (0.429) \end{matrix}$$

$$\hat{\beta}_4 = \begin{matrix} -1.017, \\ (0.365) \end{matrix} \quad \hat{\beta}_5 = \begin{matrix} -0.369, \\ (0.192) \end{matrix} \quad \hat{\beta}_6 = \begin{matrix} -0.126 \\ (0.065) \end{matrix}$$

with $\hat{\sigma}_\epsilon = 0.317$ and $Q(10) = 6.22$, where standard errors are in parenthesis below the estimates. The value $Q(d.f.)$ is the Ljung-Box statistic with $d.f. =$

11

degrees of freedom, which compared with a Chi-square distribution with 10 degrees of freedom did not show any evidence of residual autocorrelation. Then we compared the ratio (0.923-1)/0.457=-0.168 with its appropriate tabled distribution (Table B.6 for $n = 50$ in Hamilton, 1994). The corresponding 10% and 90% critical values are -1.61 and 0.91, so that H_0 is not rejected by the data and $d = 2$ becomes an appropriate choice.

Once the matrices S_2 and $\tilde{W}$ were built, we decided to fix the desirable precision in 75% and found the corresponding value $\hat{\sigma}_t^2 = 0.00171286$. Then we followed steps 3 and 4 of our recommended procedure and observed that 12 original values of $\{T(u_x^0)\}$ were corrected to obtain $\{T(u_x)\}$. Those were the values corresponding to ages: 60, 62, 64, 66, 68, 71, 79, 84, 92, 93, 94 and 97. Finally, steps 4 and 5 were followed and the Ljung-Box statistic $Q(12) = 22.94$ became significant at the 2.8% level, so that autocorrelation was still present in the residuals $\{\hat{e}_x\}$, in particular the first order autocorrelation coefficient took the value $\hat{\rho} = -0.426$.

Another round of the procedure was called for, with a higher precision value in order for the residual autocorrelation to dissapear. This must be so because more observed values will tend to be corrected for outliers. We tried then with a 90% precision and found that $\hat{\sigma}_t^2 = 0.00003879$ satisfied this criterion. Now 14 original values were corrected: the previous ones, except that of age 64, plus those of ages 75, 90 and 95. In this case the Ljung-Box statistic $Q(12) = 10.0$ did not show any evidence of autocorrelation. Similarly the skewness and (excess) kurtosis statistics (see Kendall, Stuart and Ord, 1987) took the values $Sk = -0.10$ and $Ku = -1.08$ which were reasonable for a null hypothesis that each of the true values was zero. In fact, an asymptotic test indicates to compare those values with a $N(0,1)$ distribution, thus the Normality assumption was also supported empirically. Then, the t statistic for the null hypothesis that $E(e_x) = 0$ was t=0.12 and again, there was no basis for suspicion in the contrary. Next, a visual inspection of the standarized residuals shown in Figure 2 provided support to the constant variance assumption.

<div align="center">FIGURE 2 AROUND HERE</div>

Therefore we have obtained an estimated series from a model that is valid in data. In Figure 3 we present the observed and estimated rates in the transformed scale, together with 90% confidence bands. There we can see that the assumption of true rates being integrated of order 2 makes sense, because the basic pattern is that of an adaptive straight line (i.e. an adaptive polynomial of degree 1). Further, it is also clear that the larger variability in the extremes is accounted for by wider confidence intervals. By transforming back to the original scale, we obtained the results shown in Figure 4, which is

<div align="center">12</div>

the final product of this exercise. The smoothness achieved is evident by eye, however we can also measure it numerically now. In fact, we can say that 90% of the precision (smoothness) achieved is due to the use of a dynamic structure (Assumption 3) on the true rates.

<center>FIGURES 3 AND 4 AROUND HERE</center>

We should be aware that an extremely smooth structure for the estimated rates may be undesirable for various reasons. Firstly, because it would correspond to a very rigid pattern, as can be seen by making $\sigma_t^2 = 0$ in (3.6), so that the adaptivity of the implicit polynomial is lost. Secondly, from an empirical point of view, since by making the precision get closer to 100% we will have to discard more observations through Huber's function and the residuals will show a strange behavior due to truncation. And thirdly, because $\sigma_t^2 \to 0$ will introduce numerical instabilities in the solution. In fact, we tried with 95% precision and verified the previous statements, in particular 18 observed values needed correction (rather than 14 with 90% precision) and this originated valleys and peaks in the residual pattern.

6   Final remarks

We have shown that a classical statistical interpretation of the graduation problem produces an optimal theoretical result that can be translated into a sensible working solution. Therefore, the arbitrary decisions required to graduate an observed series of rates can be put and understood in more conventional statistical terms. The key statistical arguments employed in the derivations supporting our suggested procedure, are the following. (i) Use of an appropriate transformation that helps to validate the assumptions about the random errors. (ii) Application of a robust technique that reduces the influence of outlying observations, particularly on the autocorrelation structure. (iii) The ideas of unit roots and cointegration in time series data, so that the smoothness standard (implicit in the value of d) derived from the observed data, can be assumed to be the same for the true series of rates. (iv) Measures of variability and precision associated to the estimator that allow us to make inferences and establish comparisons from the very data. (v) The idea of validating a model empirically, through a standard residual analysis.

The numerical illustration was aimed to help an interested analyst in using our methodology. We also intended to provide some insight into the performance of our procedure in practice and about the kind of results that it

<center>13</center>

can produce. Of course, that does not mean that we have obtained a completely objective solution to the graduation problem, since at the very beginning of our procedure we had to decide a desirable precision level. Nevertheless, we believe that such a decision can be made more easily with the aid of a number (through our measure of precision share) than solely on the basis of a visual inspection of a graph (as it is usually made by the practitioners). Besides, we should be well aware that our procedure requires a reasonably large series of rates (as in the numerical example) to work satisfactorily.

7 Acknowledgments

References

Box, G. E. P. and Jenkins, G. (1976). *Time Series Analysis, Forecasting and Control*, San Francisco: Holden Day.

Guerrero, V. M. and Johnson, R. A. (1982). Use of the Box-Cox transformation with binary response models. *Biometrika*, 69, 309-314.

Guerrero, V. M. (1993). Combining historical and preliminary information to obtain timely time series data. *International Journal of Forecasting*, 9, 477-485.

Hamilton, J. D. (1994). *Time Series Analysis*. New Jersey: Princeton University Press.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986) *Robust Statistics. The approach Based on Influence Functions*. New York: John Wiley and Sons.

Kendall, M. Stuart, A. and Ord, J. K. (1987). *Kendall's advanced theory of statistics*, vol I, 5th ed., London: Ch Griffin and Co.

Kimeldorf, G. S. and Jones, D. A. (1967). Bayesian graduation, *Transactions of the Society of Actuaries*, 11, 66-112.

London, D. (1985). *Graduation: The revision of estimates*. ACTEX Publications, U.S.A.

Shiu, E. S. W. (1986). A survey of graduation theory, *Proceedings of Symposia in Applied Mathematics*, 35, 67-84.

Taylor, G. (1992) A bayesian interpretation of Whittaker-Henderson graduation, *Insurance, Mathematics and Economics*, 11, 7-16.

Theil (1963) On the use of incomplete prior information in regression analysis, *Journal of the American Statistical Association*, 58, 401-414.

Verral, R. J. (1992) Graduation by dynamic regression methods, *Journal of the Institute of Actuaries*, 120, Part I, 153-170.

Verral, R. J. (1993) A state space formulation of Whittaker graduation, *Insurance, Mathematics and Economics*, 13, 7-14.

Fig 1. Mortality rates, observed and graduated by Whittaker method
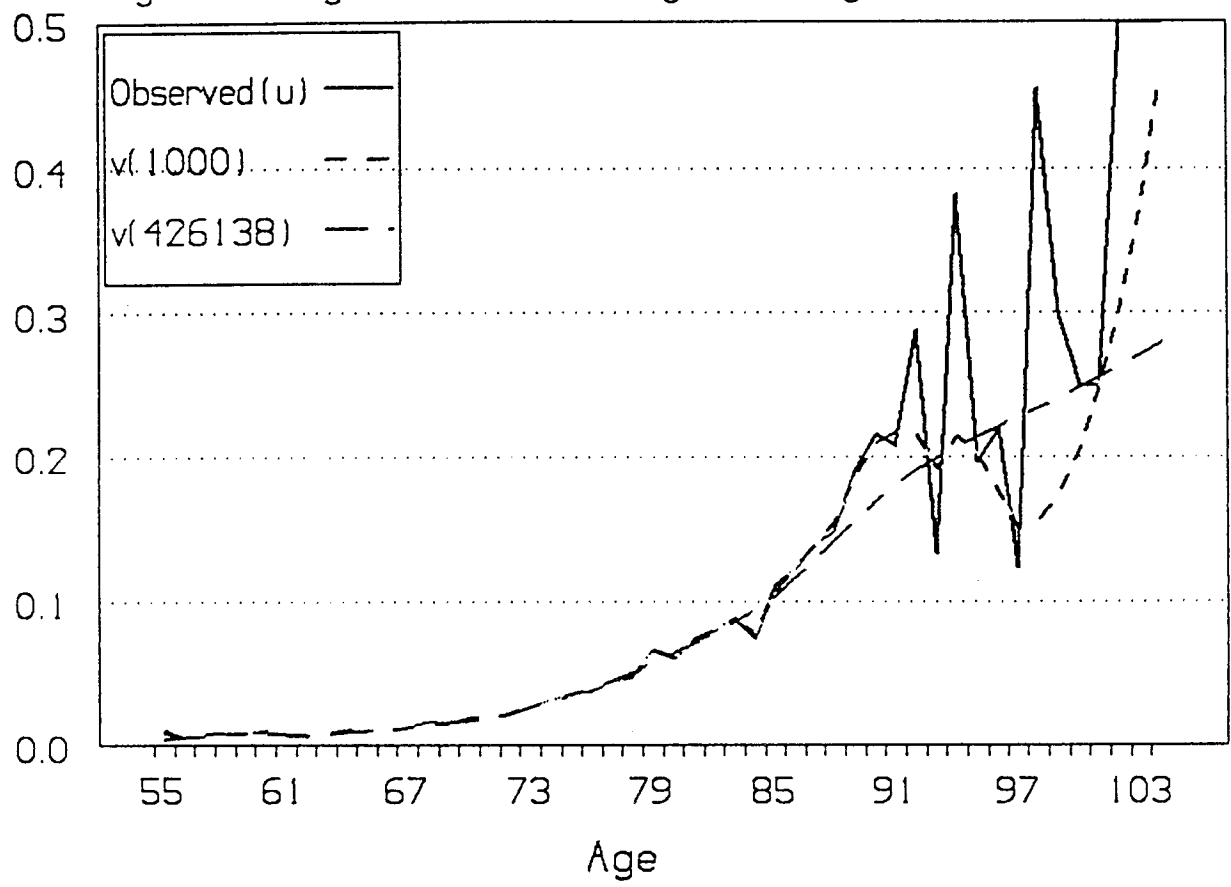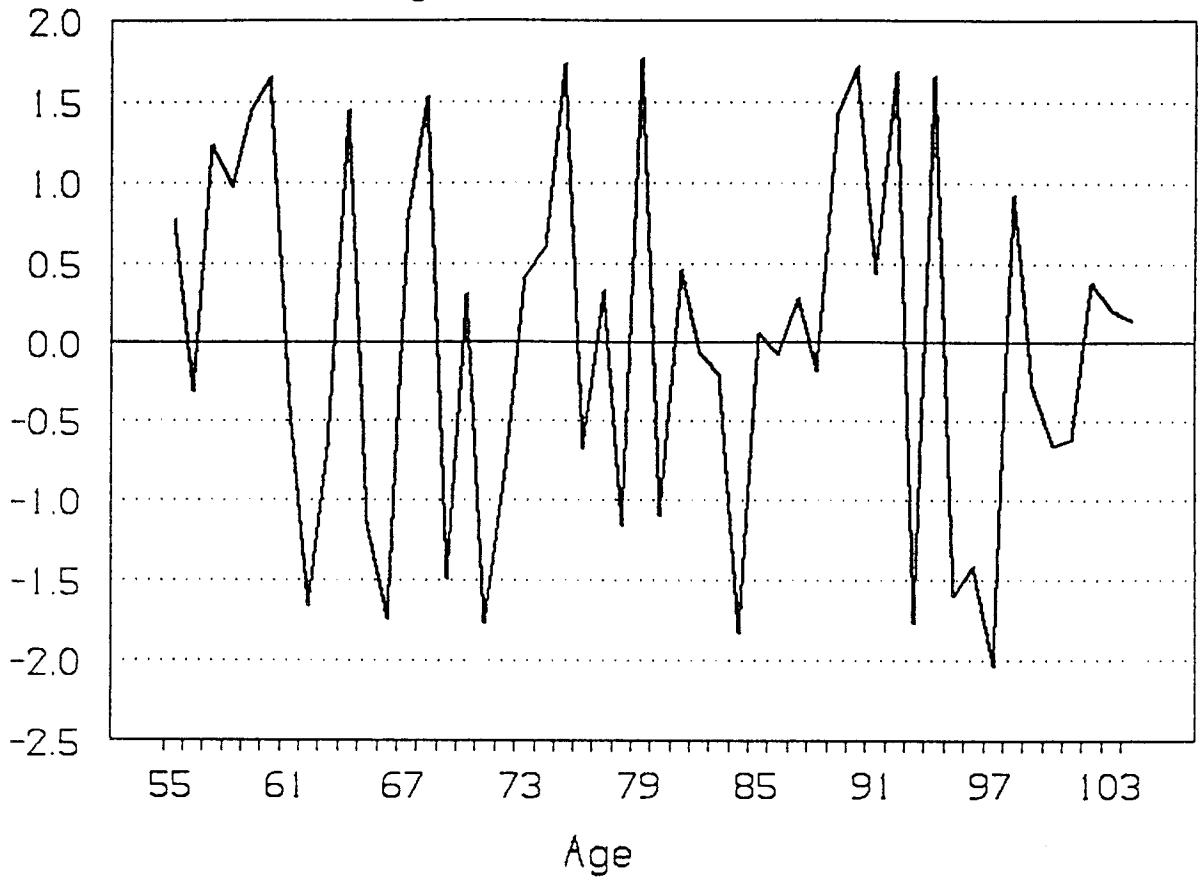
Fig 2. Standardized residuals
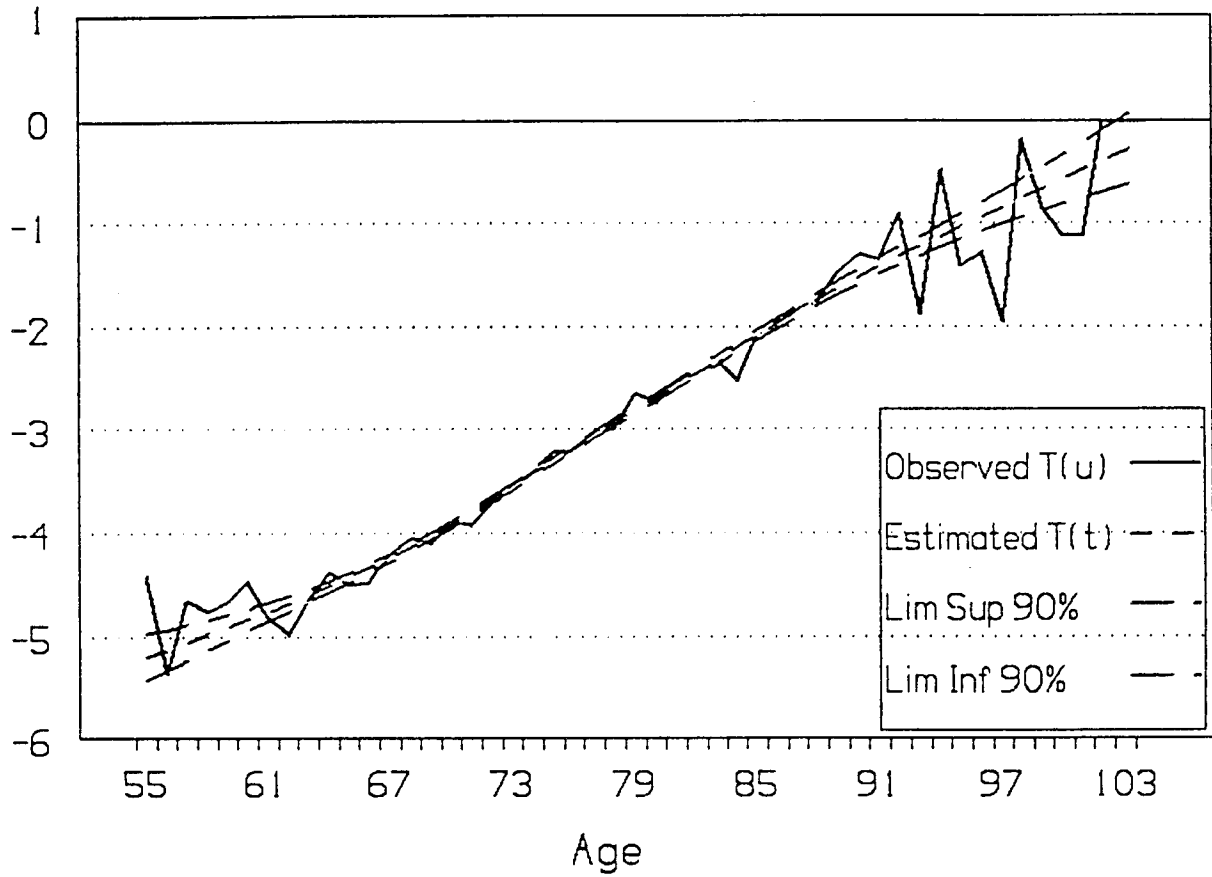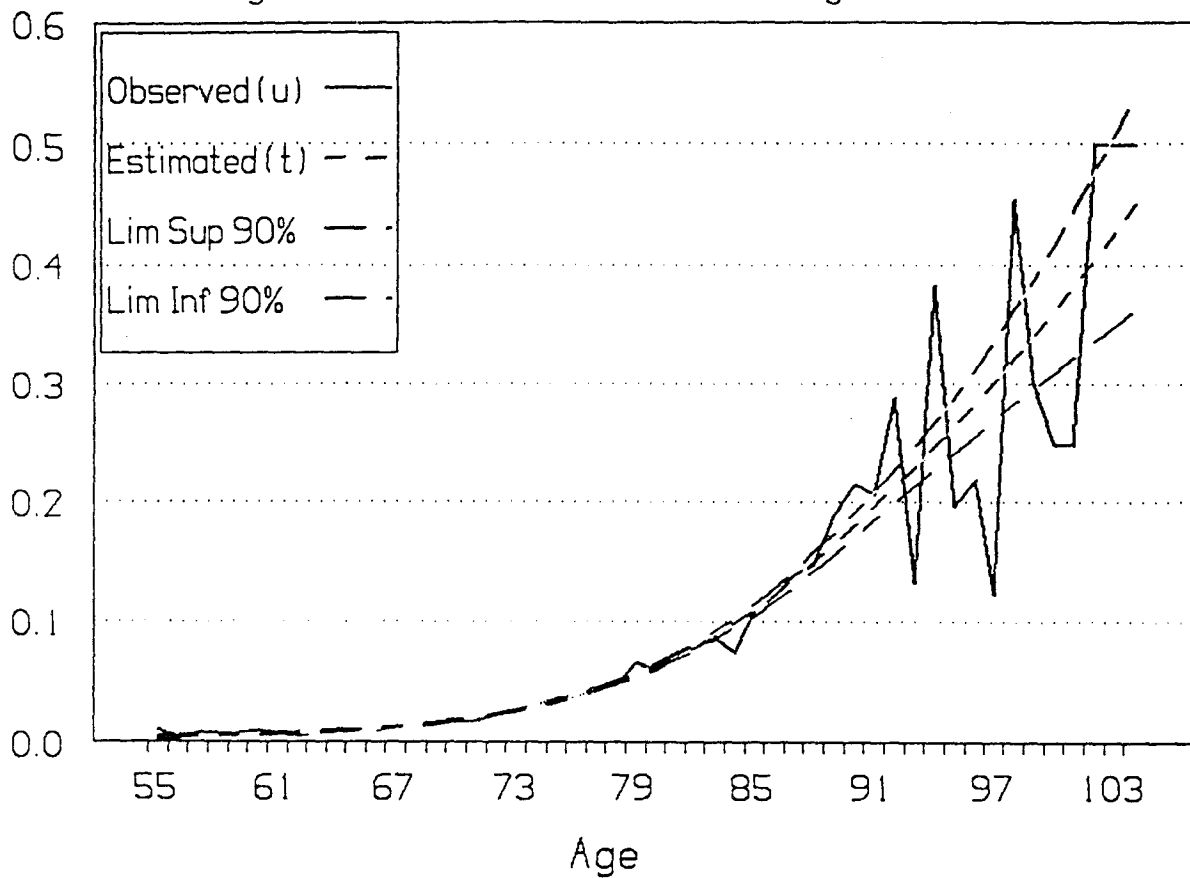
Fig 3. Observed and estimated log-odds-ratios

Fig 4. Observed and estimated rates (original scale)

Business Economics Series

97-18 (01) Margarita Samartín
            "Optimal allocation of interest rate risk"

97-23 (02) Felipe Aparicio and Javier Estrada
            "Empirical distributions of stock returns: european securities markets, 1990-95"

97-24 (03) Javier Estrada
            "Random walks and the temporal dimension of risk"

97-29 (04) Margarita Samartín
            "A model for financial intermediation and public intervention"

97-30 (05) Clara-Eugenia García
            "Competing trough marketing adoption: a comparative study of insurance companies in Belgium and Spain"

97-31 (06) Juan-Pedro Gómez and Fernando Zapatero
            "The role of institutional investors in international trading: an explanation of the home bias puzzle"

97-32 (07) Isabel Gutiérrez, Manuel Núñez Niekel and Luis R. Gómez-Mejía
            "Executive transitions, firm performance, organizational survival and the nature of the principal-agent contract"

Economics Series

97-04 (01) Iñigo Herguera and Stefan Lutz
            "Trade policy and leapfrogging"

97-05 (02) Talitha Feenstra and Noemi Padrón
            "Dynamic efficiency of environmental policy: the case of intertemporal model of emissions trading"

97-06 (03) José Luis Moraga and Noemi Padrón
            "Pollution linked to consumption: a study of policy instruments in an environmentally differentiated oligopoly"

97-07 (04) Roger Feldman, Carlos Escribano and Laura Pellisé
            "The role of government in competitive insurance markets with adverse selection"

97-09 (05) Juan José Dolado and Juan F. Jimeno
            "The causes of Spanish unemployment: a structural VAR approach"

97-10 (06) Juan José Dolado, Florentino Felgueroso and Juan F. Jimeno
            "Minimum wages, collective bargaining and wage dispersion: the Spanish case"