

GEOGRAPHIC PREMIUM RATING BY WHITTAKER SPATIAL SMOOTHING

BY

GREG TAYLOR

*Taylor Fry Consulting Actuaries, Sydney Australia
and
University of Melbourne, Australia*

ABSTRACT

Whittaker graduation is applied to the spatial smoothing of insurance data. Such data (e.g. claim frequency) form a surface over the 2-dimensional geographic domain to which they relate. Observations on this surface are subject to sampling error. They need to be smoothed spatially if a reliable estimate of the underlying surface is to be obtained.

A measure of smoothness of a surface has been defined. This has been incorporated in 2-dimensional Whittaker graduation to effect the necessary smoothing. The details of this are worked out in Section 4. The procedure is illustrated by numerical example in Section 5. The Bayesian interpretation of this form of spatial smoothing is discussed, and used to assist in the selection of the Whittaker relativity constant.

KEYWORDS

Geographic premium rating, spatial smoothing, Whittaker graduation.

1. INTRODUCTION

In certain lines of insurance business the risk varies geographically. This is typical of domestic lines, where the geographic variation may be related to directly geographic factors (e.g. traffic density, proximity to arterial roads in auto insurance) or socio-demographic factors (perhaps affecting theft rates in house insurance).

In such cases it will be desirable to estimate the geographic variation in risk premium and to price accordingly. Usually data will be available by quite fine geographic divisions, e.g. zipcodes in the US, postcodes (or sub-postcodes in the UK). The subdivisions will typically be fine enough that sampling error in each is substantial.

As a result, a mapping of sampled geographic risk takes on a rather patchy appearance. Despite this, general trends in geographic variation will often be visible. It is necessary to find a way of smoothing sampling error from one subdivision to the next in order to estimate the underlying geographic signal.

Taylor (1989) applied 2-dimensional spline functions to this problem. Boskow and Verrall (1994) provided an alternative treatment which made use of the Gibbs sampler to implement a Bayesian revision of the observations on subdivisions. The Bayesian framework recognised the magnitudes of sampling error and also incorporated the concept of smoothness over neighbouring subdivisions.

The present paper takes a rather similar approach, applying an accepted actuarial technique for compromising between smoothness and fit to data. This is Whittaker graduation, which has also been shown (Taylor, 1992; Verrall, 1993) to have a Bayesian interpretation.

2. MODEL AND NOTATION

Consider a random variable X whose mean μ is characterised by n covariates. One covariate comprises a pair of spatial (Cartesian) coordinates.

For example, X might denote claim frequency, and the coordinates might represent the centroid of a postcode region.

Although n may be any natural number, it will suffice here, and maintain brevity of notation, if the concepts are illustrated for the case $n = 3$. The extension to the general case will be obvious.

Thus, let i, j, k represent specific values of the 3 covariates, with i representing the spatial member. These values define a cell of data $\{N_{ijk}, X_{ijk}\}$, where N_{ijk} is a volume measure. In the above example it might be number of years of policy exposure.

Consistent with the notation given above,

$$E[X_{ijk}] = \mu_{ijk}. \quad (2.1)$$

Suppose that the spatial effect is separable as follows:

$$\mu_{ijk} = \nu_i \theta_{jk}, \quad (2.2)$$

for suitable parameters ν_i, θ_{jk} .

It will be supposed that the ν_i are to be estimated, but the θ_{jk} are known, perhaps by means of an earlier estimation program.

There is one degree of redundancy among the parameters ν_i, θ_{jk} . It will often be useful to set the scale of the θ_{jk} so that they are scattered about 1. If this is done, then θ_{jk} can be thought of as an adjustment multiplier to correct the quantity ν_i so that it is specific to cell ijk .

Define

$$Y_i = \frac{\sum_{jk} N_{ijk} (X_{ijk} / \theta_{jk})}{\sum_{jk} N_{ijk}}. \quad (2.3)$$

On this definition, Y_i is the summary of experience in region i but standardized for other covariates, as discussed by Brockman and Wright (1992). This adjustment to data is found in Taylor (1989) and Boskow and Verrall (1994).

By (2.1)-(2.3),

$$E[Y_i] = \nu_i, \tag{2.4}$$

showing that Y isolates the spatial effect. It will assumed that

$$V[Y_i] = \sigma^2 / N_i \tag{2.5}$$

with

$$N_i = \sum_{jk} N_{ijk}, \tag{2.6}$$

and for some suitable (though perhaps unknown) constant $\sigma^2 > 0$.

This last assumption is convenient but will often involve some degree of approximation. For example, when X denotes claim frequency with N_{ijk} X_{ijk} Poisson, one finds that

$$V[Y_i] = \phi_i \nu_i / N_i, \tag{2.7}$$

with

$$\phi_i = \sum_{jk} (N_{ijk} / N_i) \theta_{jk}^{-1}. \tag{2.8}$$

At this point, it is useful to write ν_i in the alternative form:

$$\nu_i = \nu(x_i), \tag{2.9}$$

expressing the fact that $\nu : R^2 \rightarrow R$ is a function of the spatial coordinates x .

Similarly, write $Y_i = Y(x_i)$.

3. WHITTAKER SMOOTHING

Whittaker graduation was devised by Whittaker (1923) and introduced into the actuarial literature by Henderson (1932). Since then, it has appeared in a number of standard actuarial texts, e.g. London (1985).

All of these early treatments involved smoothing a 1-dimensional sequence of observations. The generalisation to 2 or more dimensions was begun by McKay and Wilkin (1977), and a number of subsequent papers have published developments (e.g. Lowrie, 1992).

Consider points $x \in R^2$, Euclidean 2-space. The objective is to find values $W(x_i)$ which provide a smooth version of the $Y(x_i)$ and estimate $\nu(x_i)$. Define

$$D = \sum_i N_i [Y(x_i) - W(x_i)]^2 \tag{3.1}$$

which is a measure of the deviation, or error, in the observations relative to their smoothed version.

The use of $\{N_i\}$ as the set of weights in (3.1) is justified by assumption (2.5). Define

$$F = D + kS, \quad (3.2)$$

where S is a suitably chosen measure of smoothness of $W(\cdot)$, and $k(0 < k < \infty)$ is the tuning constant, or relativity constant. This constant is often chosen empirically, although Taylor (1992) and Verrall (1993) have given an analytical basis.

Whittaker smoothing consists of choosing $\{W(x_i)\}$ so as to minimise (3.2), thus achieving a compromise between error and smoothness.

The choice of S can conveniently be based on the theory of **thin plate splines** (see e.g. Green and Silverman, 1994), which uses the following penalty for lack of smoothness:

$$S = \int \left[\left(\frac{\partial^2 W}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 W}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 W}{\partial y^2} \right)^2 \right] dx dy \quad (3.3)$$

for $W = W(x, y)$ (using the (x, y) notation just for this equation) and with integration over the entire spatial region of interest.

Approximated in finite form, this is:

$$S = \sum_i S(x_i) \quad (3.4)$$

with

$$S(x) = [\Delta_{11}^2 W(x)]^2 + 2[\Delta_{12}^2 W(x)]^2 + [\Delta_{22}^2 W(x)]^2 \quad (3.5)$$

and

$$\Delta_{pq}^2 W(x) = \Delta_p [\Delta_q W(x)], \quad p, q = 1, 2 \quad (3.6)$$

$$\Delta_q W(x) = W(x + e_q) - W(x), \quad (3.7)$$

with e_1, e_2 denoting $(1, 0)$ and $(0, 1)$ respectively. Thus, Δ_q is the difference operator in the direction of the q -th coordinate axis.

4. APPLICATION

4.1. Smoothing

The basic procedure of Whittaker smoothing needs to be adapted to the situation in which the arrangement of points x at which $Y(\cdot)$ is sampled is irregular, rather than forming a lattice as in Section 3. In the case of general $\{x_i\}$, it is not clear how the differences $\Delta_{pq}^2 W(x_i)$ should be calculated.

Note that, in the lattice case, $W(y)$ is required at 6 distinct values of y to determine the 3 terms $\Delta_{pq}^2 W(x)$ at a fixed x . This reflects the fact that a general quadratic defined on 2-space, is defined by 6 parameters.

One possibility therefore would be to select 5 points “close to” a given x and fit a quadratic $Q_x(\cdot)$ to the 6 points. The values $\Delta_{pq}^2 W(x)$ for that x could then be read off from the coefficients of $Q_x(\cdot)$.

To the extent that $Q_x(\cdot)$ is merely an approximation for $W(\cdot)$ in the neighbourhood of x , there will be disagreement between the functions $Q_x(\cdot)$ defined by different x . Fitting a quadratic through precisely 6 points will cause that function to be highly sensitive to the values assumed at those points. In the same way, working with a function $f : R \rightarrow R$, fitting a 5th degree polynomial to each set of 6 consecutive values $\{f(x), f(x + 1), \dots, f(x + 5)\}$ would be liable to produce eccentric fitted functions and a high degree of conflict between them.

For this reason, it is doubtful that the most meaningful smoothness measure is obtained by the precise quadratic fitting described above. One alternative, and preferable, procedure is to fit each $Q_x(\cdot)$ by reference to a larger number of points than 6. Let h be the number of such points.

This is done in Appendix B and the smoothness measure $S(x)$ in (3.5) calculated by reference to the fitted $Q_x(\cdot)$. The calculation procedure is as follows.

Suppose that $Y(x)$ is observed at m points x_1, \dots, x_m . Let z denote the vector $[W(x_1), \dots, W(x_m)]^T$ of smoothed observations. Let $y_x^{(1)}, \dots, y_x^{(h)}$ be the subset of $\{x_i\}$ consisting of the h points closest to x , including $y_x^{(i)} = x$ and let

$$z_x = [W(y_x^{(1)}), \dots, W(y_x^{(h)})]^T.$$

For any $y \in R^2$, write $y = (y_1, y_2)$ and

$$y^\otimes = \left(\frac{1}{2}y_1^2, y_1y_2, \frac{1}{2}y_2^2, y_1, y_2, 1 \right)^T, \tag{4.1}$$

and define X_x as the $h \times 6$ matrix with $[y_x^{(i)}]^\otimes$ in its i -th row.

Define

$$A_x = (X_x^T X_x)^{-1} X_x^T, \tag{4.2}$$

and B_x as the $6 \times m$ matrix containing the h columns of A_x placed within B_x in the same positions as the components of z_x occupy in z .

Define \tilde{B}_x as the $3 \times h$ submatrix of B consisting of the latter’s first 3 rows. Then

$$S(x) = z^T \tilde{B}_x^T C \tilde{B}_x z, \tag{4.3}$$

$$S = z^T \left[\sum_i \tilde{B}_{x_i}^T C \tilde{B}_{x_i} \right] z, \tag{4.4}$$

with

$$C = \text{diag}(1, 2, 1). \tag{4.5}$$

The Whittaker criterion (3.2) may be written in the matrix notation established above.

$$F = (Y - z)^T \Lambda (Y - z) + kz^T Mz, \quad (4.6)$$

where Y is the m -vector of observations $Y(x_i)$,

$$\Lambda = \text{diag}(N_1, \dots, N_m), \quad (4.7)$$

$$M = \sum_i \tilde{B}_{x_i}^T C \tilde{B}_{x_i}. \quad (4.8)$$

Minimisation of F is carried out by differentiating (4.6) with respect to z and setting the result equal to zero. This yields the smoothed vector

$$z = (1 + k\Lambda^{-1}M)^{-1}Y. \quad (4.9)$$

The smoothed vector is equal to the unsmoothed plus a ‘‘smoothing correction’’ $k\Lambda^{-1}MY$. The greater the value chosen for the relativity constant k , the greater the correction. The correction made at any point is inversely proportional to the volume of experience at that point; the greater that volume, the less it requires smoothing.

Note that, because F is quadratic in Y and z , a scale change in Y induces the same scale change in z , provided that k is changed appropriately. This allows the useful device of replacing Y_i , defined in (2.3), by $Y_i/\bar{\nu}$, where $\bar{\nu}$ is in some sense an overall average of the ν_i .

This converts the observed values of Y_i to a scatter about 1, and (2.4) and (2.7) are replaced by $E[Y_i] = \nu_i/\bar{\nu}$ and $V[Y_i] = (\nu_i/\bar{\nu})\phi_i/N_i\bar{\nu}$.

The Bayesian interpretation of k may be obtained from Taylor (1992). Straightforward extension of the 1-dimensional reasoning given there to 2 dimensions shows that, if all second differences of z in coordinate axis directions are viewed as subject to independent priors each with variance τ^2 , then

$$k = \sigma^2/4\tau^2, \quad (4.10)$$

with σ^2 defined by (2.5).

4.2. Zoning

Consider the framework established in Section 2, in which some pricing function, such as claim frequency is being estimated by postcode. The smoothing formula (4.9) produces such estimates. In principle, it is feasible to price accordingly, i.e. postcode by postcode.

Often, however, an insurer will wish to group postcodes into convenient rating zones, or regions. For pricing purposes, the geographic effect will be taken as constant over such a zone.

A procedure for achieving this would be as follows:

1. Map the smoothed vector z to postcodes on a proper geographic map of the whole region being priced.
2. Colour code the postcodes according to the values of z_i . For example, values of z_i might range (mainly, ignoring a scatter of extreme values) from 40% for rural areas to 150% for inner city areas. In this case, different colours might be applied to the ranges $< 70\%$, $70\text{--}80\%$, $80\text{--}90\%$, etc. The colours should be spectrally ordered, e.g. red, pink, orange, yellow, etc. While the constant bandwidth illustrated above is simply to apply, it may be neater to use a multiplicative (logarithmic) scale, e.g. $91\text{--}100\%$, $100\text{--}110\%$, $110\text{--}121\%$, etc.
3. Scan the map to select zones consisting largely of the one colour, or of a small number of colours adjacent in the chromatic scheme. This requirement of chromatic homogeneity will need to be balanced against the desirability of spatially continuous connected zones.
4. Re-fit the whole model with the collection of selected zones introduced as a rating variable.

In Step 4, the model structure (2.2) is still assumed, but the spatial index i now applies to the coarser zoning determined in Step 3 instead of individual postcode.

In addition, θ_{jk} can no longer be assumed known, since these values will have been estimated with geographic effects ignored. Now that the effects are “known”, the θ_{jk} need to be reestimated, taking them into account.

For example, suppose that (2.2) may be expanded in the form:

$$\mu_{ijk} = \nu_i \theta_j \phi_k; \quad (4.11)$$

equivalently,

$$\log \mu_{ijk} = \log \nu_i + \log \theta_j + \log \phi_k. \quad (4.12)$$

Then the usual regression modelling (e.g. generalised linear) may be applied to estimate the parameters. Note that, at this stage, the numerical results obtained from the smoothing (the vector z) are discarded, and the only role played by the smoothing is the determination of zones.

An example is provided in Section 5.

4.3. Further research

The smoothing procedure described in the foregoing sections appears to work effectively most of the time. It seems well adapted to metropolitan areas, usually characterised by densely packed postcodes.

An example of cases in which it tends to break down is illustrated in schematic form in Figure 1. The polygons in the figure are a stylised representation of postcodes in a rural area. The circled numbers label the 10 postcodes. The uncircled numbers represent unsmoothed values of geographic risk, i.e. $Y(x_i)$.

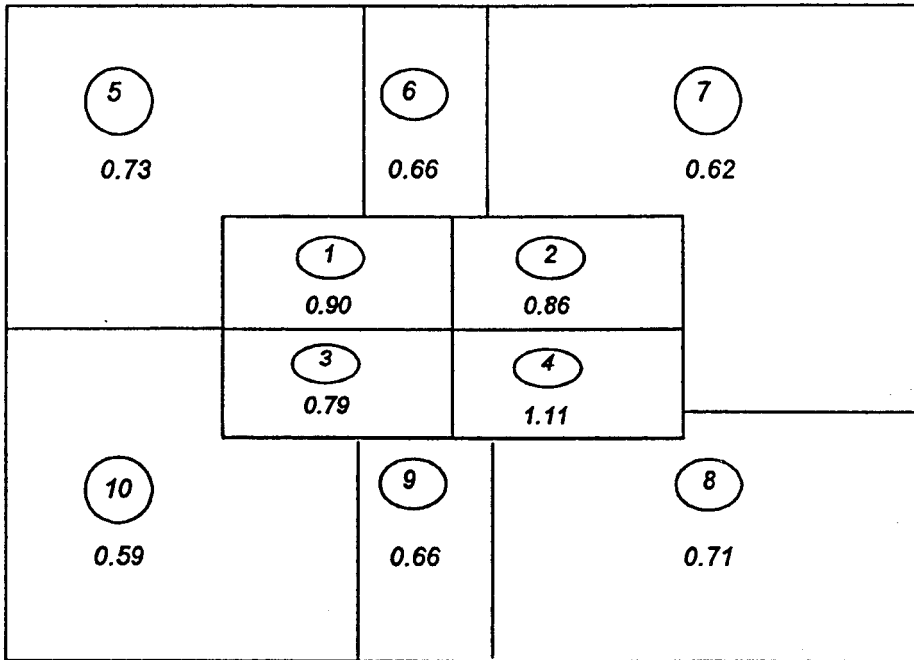


FIGURE 1: Example of geographic variation.

The small postcodes 1-4 represent a regional town, and postcodes 5-10, geographically much larger, represent surrounding rural districts.

The spatial smoothing of Section 4.1 seems poorly adapted to these circumstances. The difficulty relates to the highly localised variation of postcodes 1-4 from the surrounding trend.

According to Section 4.1, the smoothness measure for postcode 1 is calculated by reference to h postcodes. Suppose $h = 10$, and the 10 relevant postcodes are those appearing in Figure 1. In the detail of Appendix B, the smoothness measure is calculated from the curvature of the surface fitted over these 10 postcodes.

The surface needs to be mostly flat, but with a peak concentrated over a small area formed by postcodes 1-4. It is impossible to obtain this by fitting a quadratic surface, which will be much flatter than the experience of the diagram suggests.

This means that the smoothing algorithm sees postcodes 1-10 as relatively smooth before smoothing, and so applies little smoothing to them. The “smoothed” results are likely to exhibit the same “patchiness” over postcodes 1-4 as found in the unsmoothed.

Such failures are less likely to occur in metropolitan areas, since the localised extremes which cause the difficulty are less likely in these cases.

The solution to this difficulty may lie in some form of variation of h with the local topography of the $Y(x)$ surface. Such techniques might be akin to adaptive kernel smoothing of the type discussed by Bailey and Gatrell (1995, p. 87).

5. EXAMPLE

The smoothing procedure derived in Section 4.1 is applied to a particular claim frequency data set in Figures 2 to 6. These maps show smoothing produced by different choices of the relativity constant k , which increases steadily from Figure 2 to Figure 6.



FIGURE 2: No smoothing ($k = 0$).

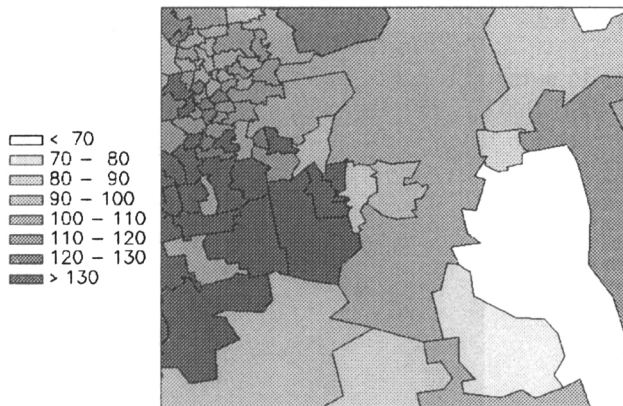


FIGURE 3: Smoothing with $k = 100$

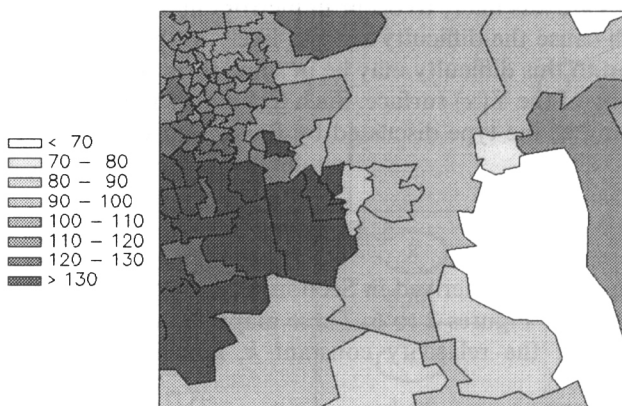


FIGURE 4: Smoothing with $k = 500$.

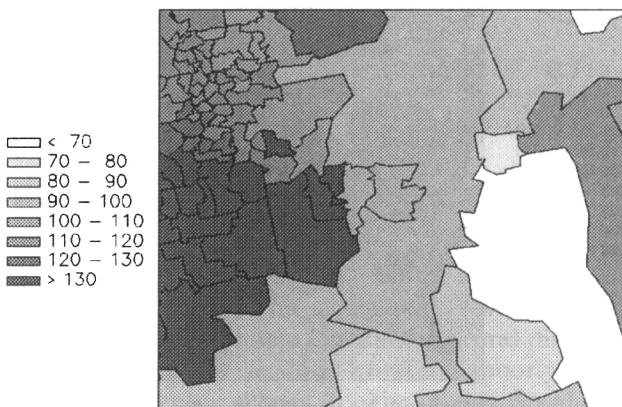


FIGURE 5: Smoothing with $k = 1000$.

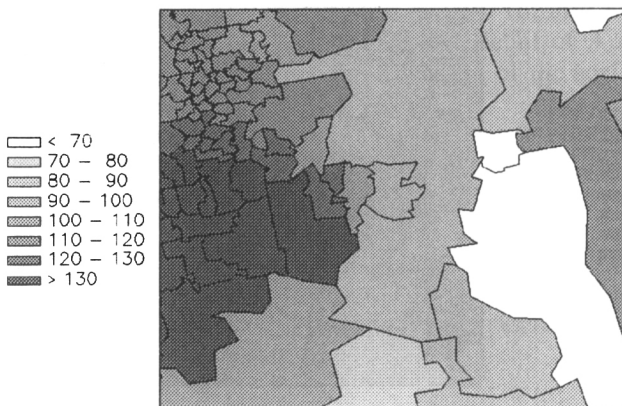


FIGURE 6: Smoothing with $k = 5000$.

Figure 2 uses $k = 0$, which reproduces the unsmoothed statistics (see (4.9)). These statistics are “residual ratios” from a regression which models all effects other than geographic. The residual ratios are defined as follows:

$$\text{Residual ratio for postcode } i = \frac{\text{number of claims observed in postcode}}{\text{model fit of this number}}$$

This is the case in which X_{ijk} represents claim frequency, but Y_i has been rescaled by the device mentioned at the end of Section 4.1. The rescaled values of Y_i are scattered about 1, and the legend in Figures 2 to 6 expresses the ratio as a percentage. For example, the band 70-80 includes residual ratios 70-80%.

The example illustrated here uses $h = 10$. The increased smoothing power resulting from increasing k is evident through the sequence of diagrams.

Equation (4.10) gives the theoretical value of k . The parameter σ^2 is defined by (2.5). For Y_i representing claim frequency, $N_i Y_i$ might be assumed Poisson, with N_i denoting exposure. This case, together with the rescaling of ν_i effected by the residual ratios, is dealt with at the end of Section 4.1, where

$$V[Y_i] = (\nu_i/\bar{\nu})\phi_i/N_i\bar{\nu}, \tag{5.1}$$

giving

$$\sigma^2 = (\nu_i/\bar{\nu})\phi_i/\bar{\nu}. \tag{5.2}$$

Now the values of $\nu_i/\bar{\nu}$ are scattered about 1, and ϕ_i will also be of the order 1 if the θ_{jk} are “centralised” in the manner suggested just prior to (2.3). Thus, a rough approximation is $\sigma^2 = 1/\bar{\nu}$, so that (4.10) gives $k = 1/4\bar{\nu}\tau^2$.

Alternatively, one might take σ^2 to be somewhat greater than this, to allow for some overdispersion relative to Poisson.

The nature of τ^2 is described at the end of Section 4.1. It is clearly more difficult to estimate, but some indication can be obtained.

For reasons of data confidentiality, an estimate of $\bar{\nu}\tau^2$ is given here, rather than separate estimates of $\bar{\nu}$ and τ^2 . Based on Figure 5, a reasonable estimation of $\bar{\nu}\tau^2$ appears to lie in the range 1/8,000 to 1/6,000, giving k in the range 1,500 to 2,000. This just fails to match $k = 1,000$ in Figure 5.

Alternatively, consider Figure 6 which suggests a value of τ^2 smaller by a factor of perhaps 2, giving k in the range 3,000 to 4,000. This fails to match $k = 5,000$ in Figure 6, but the discrepancy lies in the opposite direction from that of Figure 5.

In view of the roughness of the estimated τ^2 , it is inappropriate to regard the above calculations too literally. To the extent that they are meaningful, however, they indicate a value of k between 1,000 and 5,000. This suggests that Figure 6 may be over-smoothed, while Figures 3 and 4 are perhaps under-smoothed.

ACKNOWLEDGEMENT

The author acknowledges the assistance of Steven Lim, who programmed the smoothing algorithm developed in this paper, and linked it to the mapping software used to produce Figures 2 to 6. The assistance of Dr. Richard Verrall, who provided several references to multi-dimensional smoothing procedures, is also acknowledged.

APPENDIX A

FITTING A QUADRATIC DEFINED ON 2-SPACE TO h OBSERVATIONS

Consider a quadratic $Q : R^2 \rightarrow R$. Let $y^{(1)}, \dots, y^{(h)} \in R^2$ and $z^{(1)}, \dots, z^{(h)}$ be "observations":

$$z^{(i)} = Q(y^{(i)}) + \text{error}. \quad (\text{A.1})$$

If the quadratic is written out explicitly, it is

$$Q(y) = q_{20}y_1^2 + q_{11}y_1y_2 + q_{02}y_2^2 + q_1y_1 + q_2y_2 + q_0, \quad (\text{A.2})$$

where $y = (y_1, y_2)^T$. Write this as:

$$Q(y) = q^T y^\otimes, \quad (\text{A.3})$$

with

$$q = (q_{20}, q_{11}, q_{02}, q_1, q_2, q_0)^T, \quad (\text{A.4})$$

$$y^\otimes = (y_1^2, y_1y_2, y_2^2, y_1, y_2, 1)^T. \quad (\text{A.5})$$

By (A.1) and (A.3),

$$z^{(i)} = q^T [y^{(i)}]^\otimes + \text{error}, \quad i = 1, \dots, h. \quad (\text{A.6})$$

OLS regression of the $z^{(i)}$ on the $[y^{(i)}]^\otimes$ yields the following estimate \hat{q} of q :

$$\hat{q} = Az, \quad (\text{A.7})$$

with

$$A = (X^T X)^{-1} X^T, \quad (\text{A.8})$$

where

$$z = (z^{(1)}, \dots, z^{(h)})^T, \quad (\text{A.9})$$

and X is the $h \times 6$ matrix with $[y_i^\otimes]^T$ as i -th row.

APPENDIX B
CALCULATION OF THE SMOOTHNESS MEASURE

Section 4.1 requires that a local approximation $Q_x(\cdot)$ be fitted to $W(\cdot)$ by reference to the values of $W(y)$ at h points y .

The fitting can be carried out by OLS regression. Write

$$Q_x(y) = q_x^T y^\otimes \tag{B.1}$$

where $y = (y_1, y_2)^T$, and

$$y^\otimes = \left(\frac{1}{2}y_1^2, y_1y_2, \frac{1}{2}y_2^2, y_1, y_2, 1\right)^T, \tag{B.2}$$

and q_x is the corresponding vector of coefficients.

Let $y_x^{(1)}, \dots, y_x^{(h)}$ be the h points closest to x , including $y_x^{(1)} = x$, for which $Y(\cdot)$ is sampled, and let

$$z_x = \left[W(y_x^{(1)}), \dots, W(y_x^{(h)})\right]^T. \tag{B.3}$$

Appendix A shows that

$$q_x = \begin{matrix} A_x & z_x \\ 6 \times 1 & \begin{matrix} 6 \times h & h \times 1 \end{matrix} \end{matrix}, \tag{B.4}$$

where

$$A_x = (X_x^T X_x)^{-1} X_x^T, \tag{B.5}$$

and X_x is the $h \times 6$ matrix with $\left[y_x^{(i)}\right]^{\otimes T}$ as i -th row.

The result (B.4), which is expressed in terms of the ‘‘local’’ set of smoothed values z_x , needs to be expressed in terms of the global set $z = [W(x_1), \dots, W(x_m)]^T$ corresponding to the whole set of observations $Y(x_i), i = 1, \dots, m$.

This is done by rewriting (B.4) as:

$$q_x = \begin{matrix} B_x & z \\ 6 \times 1 & \begin{matrix} 6 \times m & m \times 1 \end{matrix} \end{matrix}, \tag{B.6}$$

where B_x is the $6 \times m$ matrix containing the h columns of A_x placed so as to reference h components of z_x as components of z , and zeros elsewhere.

The required differences $\Delta_{pq}^2 W(x)$ can now be approximated by the corresponding differences of $Q_x(x)$, which are given by the first 3 components of q_x . The relevant part of (B.6) is therefore

$$\tilde{q}_x = \begin{matrix} \tilde{B}_x & z \\ 3 \times 1 & \begin{matrix} 3 \times m & m \times 1 \end{matrix} \end{matrix}, \tag{B.7}$$

where the tilde indicates the operation ‘‘take the first 3 rows of’’.

It is now possible to express $S(x)$ from (3.5) as a quadratic form in \tilde{q}_x :

$$S(x) = \tilde{q}_x^T C \tilde{q}_x \tag{B.8}$$

with

$$C = \text{diag}(1, 2, 1). \quad (\text{B.9})$$

By (B.7) and (B.8),

$$S(x) = z^T \tilde{B}_x^T C \tilde{B}_x z, \quad (\text{B.10})$$

and finally, by (3.4),

$$S = z^T \left[\sum_i \tilde{B}_{x_i} C \tilde{B}_{x_i} \right] z. \quad (\text{B.11})$$

REFERENCES

- BAILEY, T.C. and GATRELL, A.C. (1995) *Interactive spatial data analysis*. Longman, England.
- BOSKOW, M. and VERRALL, R.J. (1994) Premium rating by geographic area using spatial models. *ASTIN Bulletin* **24**, 131-143.
- BROCKMAN, M.J. and WRIGHT, T.S. (1992) Statistical motor rating: making effective use of your data. In *Journal of the Institute of Actuaries* **119**, 457-526.
- GREEN, P.J. and SILVERMAN, B.W. (1994) *Non-parametric regression and generalised linear models: a roughness penalty approach*. Chapman and Hall, London, New York.
- HENDERSON, R. (1924) A new method of graduation. In *Transactions of the Actuarial Society of America* **25**, 29-40.
- LONDON, R. (1985) *Graduation: the revision of estimates*. ACTEX, Winsted and Abington, CT.
- LOWRIE, W.B. (1992) Multidimensional Whittaker-Henderson graduation with constraints and mixed differences. In *Transactions of the Society of Actuaries* **45**, 215-255.
- MCKAY, S.F. and WILKIN, J.C. (1977) Derivation of a two-dimensional Whittaker-Henderson type B graduation formula. Appendix to *Experience of Disabled-Worker Benefits Under OASDI, 1965-74*. Actuarial Study No. 74, U.S. Department of Health, Education, and Welfare.
- TAYLOR, G.C. (1989) Use of spline functions for premium rating by geographic area. *ASTIN Bulletin* **19**, 91-122.
- TAYLOR, G.C. (1992) A Bayesian interpretation of Whittaker-Henderson graduation. In *Insurance: mathematics and economics* **11**, 7-16.
- VERRALL, R.J. (1993) A state space formulation of Whittaker graduation, with extensions. In *Insurance: mathematics and economics* **13**, 7-14.
- WHITTAKER, E.T. (1923) On a new method of graduation. In *Proceedings of the Edinburgh Mathematical Society* **41**, 63-75.

GREG TAYLOR

Taylor Fry Consulting Actuaries
 Level 4, 5 Elizabeth Street
 Sydney NSW 2000
 Australia

Professorial Associate, Centre for Actuarial Studies
 Faculty of Economics and Commerce
 University of Melbourne
 Parkville VIC 3052
 Australia