LOCAL POLYNOMIAL REGRESSION AND THE HENDERSON FILTER

Local Polynomial Regression

There is a way of looking at moving averages and kernel smoothing which will prove fruitful hereafter. Consider the quadratic function

$$S(t) = \sum_{j=-m}^{m} \psi_j (y_{t-j} - x_t)^2. \tag{1}$$

The minimising value $x_t = \sum_j \psi_j y_{t-j}$ is just a weighted average of data values in the vicinity of y_t. The weights of this average need to be determined. An indirect way of doing so is to seek to determine the smoothed value x_t via a local polynomial regression. The polynomial is fitted to the set of data points $y_{t-j}; j = 0, \pm 1, \ldots, \pm m$ in the vicinity of y_t that are comprised by the weighted average. If the polynomial is denoted by $\gamma(j) = \gamma_0 + \gamma_1 j + \cdots + \gamma_p j^p$, then the smoothed value will be provided by the ordinate at $j = 0$, which is to say that $x_t = \gamma_0$.

The polynomial may be fitted by minimising a weighted sum of squares of the deviations of the local data from the polynomial:

$$S(t) = \sum_{j=-m}^{m} \lambda_j \{y_{t+j} - \gamma(j)\}^2. \tag{2}$$

Then, the estimates of the polynomial coefficients will be linear functions of these data values. In particular, the minimisation of $S(t)$ will determine a set of coefficients $\{\psi_j; j = 0, \pm 1, \ldots, \pm m\}$ such that $\gamma_0 = \sum \psi_j y_{t-j}$.

To examine these possibilities in more detail, let us define

$$J = \begin{bmatrix} 1 & -m & m^2 & \ldots & (-m)^p \\ 1 & 1-m & (m-1)^2 & \ldots & (1-m)^p \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & m-1 & (m-1)^2 & \ldots & (m-1)^p \\ 1 & m & m^2 & \ldots & m^p \end{bmatrix}, \tag{3}$$

which is the matrix of a basis for the local polynomial, together with the diagonal weighting matrix

$$\Lambda = \mathrm{diag}\{\lambda_{-m}, \lambda_{1-m}, \ldots, \lambda_{m-1}, \lambda_m\}. \tag{4}$$

Then the vector $\gamma = [\gamma_0, \gamma_1, \ldots, \gamma_p]'$ of polynomial coefficients is obtained as the solution to the following normal equations of a local polynomial regression:

$$J'\Lambda J\gamma = J'\Lambda y, \tag{5}$$

wherein $y = [y_{t-m}, y_{t+1-m}, \ldots, y_{t+m-1}, y_{t+m}]'$. The smoothed value to replace y_t is

$$\begin{aligned}\gamma_0 = e_0\gamma &= e_0(J'\Lambda J)^{-1}J'\Lambda y \\ &= \psi' y,\end{aligned} \tag{6}$$

where $e_0 = [1, 0, \ldots, 0]'$; and it is manifest that the filter weights in $\psi = [\psi_{-m}, \ldots, \psi_m]'$ do not vary as the filter passes through the sample. Also, it can be seen that $\psi' J = e_0(J'\Lambda J)^{-1}J'\Lambda J = e_0$, which is to say that

$$\sum_{j=-m}^{m} \psi_j = 1, \ \sum_{j=-m}^{m} j\psi_j = 0, \ \sum_{j=-m}^{m} j^2\psi_j = 0, \ldots, \ \sum_{j=-m}^{m} j^p\psi_j = 0. \tag{7}$$

These conditions are necessary and sufficient to ensure that $\sum_{j=-m}^{m} \psi_j\gamma(j) = \gamma_0$, which implies that the filter transmits, without alteration, any polynomial function of degree p.

The technique of filtering via local polynomial regression becomes fully specified only when the regression weights within Λ are determined. The matter is dealt with in the following section.

The Henderson Filter

The requirement of Henderson (1916) was for a symmetric filter that would transmit a cubic polynomial time trend without distortion. It was also required that the filtered series should be as smooth as possible.

Consider the normal equations (5) in the case where $p = 3$. The generic element in the rth row and kth column of the matrix $J'\Lambda J$ is $\sum_{j=-m}^{m} \lambda_j j^{r+k} = s_{r+k}$. The filter will be symmetric if and only if the regression weights are symmetric such that $\lambda_j = \lambda_{-j}$ and, under these conditions, it follows that $s_{r+k} = 0$ if $r + k$ is odd. Therefore, the normal equations take the form of

$$\begin{bmatrix} s_0 & 0 & s_2 & 0 \\ 0 & s_2 & 0 & s_4 \\ s_2 & 0 & s_4 & 0 \\ 0 & s_4 & 0 & s_6 \end{bmatrix} \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} = \begin{bmatrix} \sum \lambda_j y_{t-j} \\ \sum j\lambda_j y_{t-j} \\ \sum j^2\lambda_j y_{t-j} \\ \sum j^3\lambda_j y_{t-j} \end{bmatrix}. \tag{8}$$

Only the first and the third of these equations are involved in the determination of γ_0 via

$$\begin{bmatrix} \gamma_0 \\ \gamma_2 \end{bmatrix} = \begin{bmatrix} s_0 & s_2 \\ s_2 & s_4 \end{bmatrix}^{-1} \begin{bmatrix} \sum \lambda_j y_{t-j} \\ \sum j^2\lambda_j y_{t-j} \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} \begin{bmatrix} \sum \lambda_j y_{t-j} \\ \sum j^2\lambda_j y_{t-j} \end{bmatrix}. \tag{9}$$

Thus

$$\gamma_0 = \sum_{j=-m}^{m} \psi_j y_{t-j} = \sum_{j=-m}^{m} (a + bj^2)\lambda_j y_{t-j}, \tag{10}$$

where

$$as_0 + bs_2 = 1 \qquad \text{and} \qquad as_2 + bs_4 = 0. \tag{11}$$

It is now a matter of determining the filter coefficients $\psi_j = (a + bj^2)\lambda_j$ in accordance with the smoothness criterion.

The criterion adopted by Henderson was that the variance of the third differences of the filtered sequence should be at a minimum. The criterion requires that process generating sequence should be specified sufficiently for the variance to be defined. An appropriate model is one in which the data are generated by a cubic polynomial with added white noise: $y(t) = \beta(t) + \varepsilon(t)$.

The (forward) difference operator $\Delta = F - 1$ has the effect that $\Delta y(t) = y(t+1) - y(t)$. Also, the third difference of a cubic polynomial is some constant c. Therefore, $\Delta^3 y(t) = c + \Delta^3\varepsilon(t)$; and, if $V(\varepsilon_t) = \sigma^2$, it follow that

$$V\{\Delta x(t)\} = \sum_j \{\Delta^3\psi(j)\}\sigma^2, \tag{12}$$

where $\psi(j)$ denotes the indefinite extension of the sequence of filter coefficients, formed by supplementing them with the set of zero-valued elements $\{\psi_{m+j} = 0; j = \pm 1, \pm 2, \ldots\}$.

The resulting criterion is

$$\text{Minimise} \quad \sum_j \Delta^3\psi(j) \qquad \text{subject to} \quad \sum_j \psi(j) = 1, \quad \sum_j j^2\psi(j) = 0. \tag{13}$$

The two side conditions are from (7). The remaining conditions $\sum_j j\psi(j) = \sum_j j^3\psi(j) = 0$ are satisfied automatically in consequence of the symmetry about ψ_0 of the sequence $\psi(j)$. The constrained minimisation, which can be affected using Lagragean multipliers, indicates that

$$\Delta^6\psi(j-3) = a + bj^2, \quad \text{for} \quad j = 0, \pm 1, \ldots, \pm m. \tag{14}$$

This implies that the filter coefficients are the ordinates of a polynomial in j of degree 8, namely $\psi(j) = \delta(j)(a + bj^2)$, of which the 6th difference is the quadratic function $a + bj^2$. For condition of (14) to be satisfied, it is necessary that $\psi(j)$ should be specified for the additional values of $j = \pm(m+1), \pm(m+2), \pm(m+2)$. Here the ordinates are all zeros. It follows that the polynomial, which must have zeros at these six points, must take the form of

$$\psi(j) = \{(m+1)^2 - j^2\}\{(m+2)^2 - j^2\}\{(m+3)^2 - j^2\}(a + bj^2). \tag{15}$$

There are only two remaining parameters to be determined, which are a and b. They are determined via the conditions of (11).

Kenney and Durbin (1982) have found that

$$\psi_j \propto \{(m+1)^2 - j^2\}\{(m+2)^2 - j^2\}\{(m+3)^2 - j^2\}(3(m+2)^2 - 16 - 11j^2), \tag{16}$$

where the constant of proportionality is chosen to ensure that the coefficients sum to unity.

References

Henderson, R.,(1916), Note on Graduation by Adjusted Average, *Transactions of the Actuarial Society of America,* 17, 43–48

Kenny, P.B., and J. Durbin, (1982), Local Trend Estimation and Seasonal Adjustment of Economic and Social Time Series, *Journal of the Royal Statistical Society,* Series A (General), 145, 1–41.