A. The Linear first order model

For the polynomial model of order 1, the (nonlinear) least squares problem is

\[
\{\hat{q}, \hat{\beta}_1\} = \arg \min_{q, \beta_1} \sum_{i=1}^{n} (y_i - \beta_1(q_{s_i} - d_i))^2
\]

\[
= \arg \min_{r, \beta_1} \sum_{i=1}^{n} (y_i - r_{s_i} + \beta_1 d_i))^2,
\]

(1)

where we introduced the new variables \(r_i = \beta_1 q_i\), \(i = 1, \ldots, N\). Write now the \(i\)th "observation" vector \(\psi_i\), having all zero entries \(\psi_i(j) = 0\), except the two entries: \(\psi_i(s_i) = 1\) and \(\psi_i(N+1) = -d_i\). The vector of the unknowns, \(\eta\), has \(N+1\) entries as follows: \(\eta(i) = r_i, i = 1, \ldots, N\) and \(\eta(N+1) = \beta_1\). The linear least squares problem

\[
\min_{\eta} \sum_{i=1}^{n} (y_i - \psi_i^T \eta)^2
\]

(2)
gives the solution

\[
\hat{\eta} = \left( \sum_{i=1}^{n} \psi_i \psi_i^T \right)^{-1} \sum_{i=1}^{n} \psi_i y_i,
\]

(3)

from which we find easily the estimates minimizing (1) as \(\hat{q}(i) = \frac{\hat{\eta}(i)}{\hat{\eta}(N+1)}\) and

\[
\hat{q}(i) - \hat{q}(j) = \frac{\hat{\eta}(i) - \hat{\eta}(j)}{\hat{\eta}(N+1)}.
\]

(4)
Lemma 0.1: If the data \( y_i \) were generated by a general polynomial of order 1 as follows: \( y_i = \bar{\beta}_0 + \bar{\beta}_1(q_i - d_i) + e_i \) with \( e_i \sim N(0, \sigma^2) \), then

(a) The estimate \( \hat{\beta}_1 = \hat{\eta}(N + 1) \) given by (3) has the Gaussian distribution \( \hat{\beta}_1 \sim N(\bar{\beta}_1, \sigma^2_{\hat{\beta}_1}) \), where \( \sigma^2_{\hat{\beta}_1} = \frac{12\sigma^2}{\kappa r(m^2 - 1)} \), and the estimate \( \hat{\eta}(i) - \hat{\eta}(j) \sim N(\bar{\beta}_1(q_i - q_j), \sigma^2_{\hat{\eta}}) \), where \( \sigma^2_{\hat{\eta}} = \frac{2\sigma^2}{mK} \).

(b) The estimate \( \vartheta = \hat{q}_i - \hat{q}_j \) of the difference \( \bar{q}_i - \bar{q}_j \) for every pair \((i, j)\) given by (4) has a cumulative distribution function, \( D(\vartheta) \), which can be approximated by

\[
D(\vartheta) \rightarrow \Phi\left\{ \frac{\vartheta - (\bar{q}_i - \bar{q}_j)}{\sqrt{\frac{2\sigma^2}{\beta_1} + \frac{\sigma^2_{\hat{\eta}}}{\beta_1^2}}} \right\}
\]

when the coefficient of variation, \( \frac{\sigma^2_{\hat{\eta}}}{\beta_1^2} \), becomes negligible. Here \( \Phi(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \) is the normal cumulative distribution function with zero mean and unit variance.

Proof

Write \( \bar{r}_i = \bar{\beta}_1(q_i + \frac{\hat{\beta}_1}{\beta_1}) \), \( i = 1, \ldots, N \), and introduce the vector \( \bar{\eta} \) having \( N + 1 \) entries as follows \( \bar{\eta}(i) = \bar{r}_i, i = 1, \ldots, N \) and \( \bar{\eta}(N + 1) = \bar{\beta}_1 \). The generated data can be written as \( y_i = \psi_i^T \bar{\eta} + e_i \), and the estimate (3) becomes

\[
\begin{align*}
\hat{\eta} & = \left( \sum_{i=1}^{n} \psi_i \psi_i^T \right)^{-1} \sum_{i=1}^{n} \psi_i (\psi_i^T \bar{\eta} + e_i) \\
& = \bar{\eta} + \left( \sum_{i=1}^{n} \psi_i \psi_i^T \right)^{-1} \sum_{i=1}^{n} \psi_i e_i.
\end{align*}
\]

Thus \( \hat{\eta} \sim N(\bar{\eta}, \Psi) \), where \( \Psi = \bar{\sigma}^2 (\sum_{i=1}^{n} \psi_i \psi_i^T)^{-1} = \bar{\sigma}^2 \Upsilon^{-1} \). The entries of the matrix \( \Upsilon \) can be evaluated easily: for \( 1 \leq j, k \leq N \) we have

\[
\begin{align*}
\Upsilon(j, k) & = \sum_{i=1}^{n} \psi_i(j) \psi_i(k) \\
& = \sum_{i|s_i = j}^{n} \psi_i(j) \psi_i(k) \delta_{jk} = mK \delta_{jk}, \\
\Upsilon(j, N + 1) & = \sum_{i|s_i = j}^{n} -d_i = -K \sum_{l=0}^{m-1} l \\
& = -K, m(m - 1)/2, \\
\Upsilon(N + 1, N + 1) & = \sum_{i=1}^{n} d_i^2 = K \sum_{l=0}^{m-1} l^2 \\
& = K, m(m - 1)(2m - 1)/6.
\end{align*}
\]
With the notations $e_1 = mK_r, e_2 = -K_r(m-1)/2, e_3 = K_r(m-1)(2m-1)/6, 1_N$ for the vector of length $n$ consisting of all the elements 1, and $I_N$ for the unit matrix of size $N \times N$, we see that the matrix $\Upsilon$ has the block structure

$$\Upsilon = \begin{bmatrix} e_1 I_N & e_2 1_N \\ e_2 1_N^T & e_3 \end{bmatrix}. \tag{8}$$

By the inverse lemma of partitioned matrices the covariance matrix of $\hat{\eta}$ is

$$\Psi = \sigma^2 \begin{bmatrix} \frac{1}{e_1} \left( I_N + \frac{e_2}{e_3 e_1 - e_2} 1_N 1_N^T \right) & \frac{-e_2}{e_3 e_1 - e_2} 1_N \\ \frac{-e_2}{e_3 e_1 - e_2} 1_N & \frac{e_3}{e_3 e_1 - e_2} \end{bmatrix} \begin{bmatrix} - \frac{1}{K_r} \left( I_N + \frac{3(m-1)}{m+1} 1_N 1_N^T \right) \frac{6}{K_r m(m+1)} 1_N \\ \frac{6}{K_r m(m+1)} 1_N & \frac{12}{K_r m(m^2-1)} \end{bmatrix}. \tag{9}$$

It is obvious now that $\hat{\beta}_1 = \hat{\eta}(N+1) \sim N(\beta_1, \frac{12\sigma^2}{K_r m(m^2-1)})$. Write the elements in the matrix $\Psi$ as $\Psi = \sigma^2 \begin{bmatrix} e_4 I_N + e_5 1_N 1_N^T e_6 1_N \\ e_6 1_N^T e_7 \end{bmatrix}$. Fix two positive integers $i, j \leq N$ and consider the matrix $A = \begin{bmatrix} b_1^T \\ b_2^T \end{bmatrix}$ with the $(N+1)$-vector $b_1$ having all zero entries except $b_1(i) = b_1(j) = 1$ and the $(N+1)$-vector $b_2$ having all zero entries except $b_2(N+1) = 1$. Then the vector $A \hat{\eta} = \begin{bmatrix} \hat{\eta}(i) - \hat{\eta}(j) \\ \hat{\beta}_1 \end{bmatrix}$ is normally distributed with the mean $\begin{bmatrix} \hat{\eta}(i) - \hat{\eta}(j) \\ \hat{\beta}_1 \end{bmatrix}$ and the covariance matrix

$$A \Psi A^T = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \sigma^2 \begin{bmatrix} e_4 I_2 + e_5 1_2 1_2^T e_6 1_2 \\ e_6 1_2^T e_7 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \end{bmatrix} \sigma^2 \begin{bmatrix} 2e_4 \\ 0 \\ 0 \\ e_7 \end{bmatrix}. \tag{9}$$

Thus $\hat{\eta}(i) - \hat{\eta}(j)$ is not correlated with $\hat{\beta}_1$, and $\hat{\eta}(i) - \hat{\eta}(j) \sim N(\beta_1(q_i - q_j), \frac{2\sigma^2}{mK_r})$. Now we can apply directly the result from [1] to find (5). q.e.d.

**APPENDIX: FISHER INFORMATION MATRIX FOR HETEROSEDASTIC MODELS**

The Fisher information matrix has the general entry

$$J_{j,k}(\theta) = - \sum_{i=1}^{n} E \left\{ \frac{\partial^2 \ell(y_i, \theta)}{\partial \theta_j \partial \theta_k} \right\}. \tag{10}$$
\[ \ell(y_i, \theta) = \text{const} - \frac{1}{2} \log \tau(i, \theta) - \frac{(y - \mu(i, \theta))^2}{2\tau(i, \theta)}. \]  

The partial derivatives are

\[ \frac{\partial \ell(y_i, \theta)}{\partial \theta_j} = -\frac{1}{2} \frac{\partial \tau(i, \theta)}{\partial \theta_j} \tau(i, \theta) + \frac{1}{2} \frac{\partial \mu(i, \theta)}{\partial \theta_j} (y - \mu(i, \theta)) \]  

\[ + \frac{1}{2} \frac{\partial \mu(i, \theta)}{\partial \theta_j} (y - \mu(i, \theta)) \]  

\[ \frac{\partial \mu(i, \theta)}{\partial \theta_j} = \frac{\partial g(q_{\xi_i} - d_i, \beta)}{\partial \theta_j}. \]  

\[ \frac{\partial^2 \ell(y_i, \theta)}{\partial \theta_j \partial \theta_k} = \frac{1}{\tau(i, \theta)^2} \left( \frac{\partial^2 \mu(i, \theta)}{\partial \theta_j \partial \theta_k} (y - \mu(i, \theta)) - \frac{\partial \mu(i, \theta)}{\partial \theta_j} \frac{\partial \mu(i, \theta)}{\partial \theta_k} (y - \mu(i, \theta)) \right) \]  

\[ + \frac{\partial^2 \tau(i, \theta)(y - \mu(i, \theta))^2}{\partial \theta_j \partial \theta_k} - \frac{\partial \tau(i, \theta)}{\partial \theta_j} \frac{\partial \mu(i, \theta)}{\partial \theta_k} (y - \mu(i, \theta)) - \frac{\partial \tau(i, \theta)}{\partial \theta_j} \frac{\partial \tau(i, \theta)}{\partial \theta_k} (y - \mu(i, \theta))^2 \]  

so that

\[ J_{j,k} = -\sum_{i=1}^{n} E \left\{ \frac{\partial^2 \ell(y_i, \theta)}{\partial \theta_j \partial \theta_k} \right\} \]  

\[ = \sum_{i=1}^{n} + \frac{1}{2} \frac{\partial \tau(i, \theta)}{\partial \theta_j} \frac{\partial \tau(i, \theta)}{\partial \theta_k} \tau(i, \theta)^2 + \frac{\partial \mu(i, \theta)}{\partial \theta_j} \frac{\partial \mu(i, \theta)}{\partial \theta_k} \frac{1}{\tau(i, \theta)}. \]  

Since \( \mu(i, \theta) = g(q_{\xi_i} - d_i, \beta) \) and \( \tau(i, \theta) = \sigma_\theta^2 g(q_{\xi_i} - d_i, \beta)^{2\alpha} \) the derivatives involved can be computed as follows (we take \( \alpha \neq 0 \)):

\[ \frac{\partial g(q_{\xi_i} - d_i, \beta)}{\partial \theta_j} = \begin{cases} \delta(s_i = j)g'_{\beta_{\xi_i}}(q_{\xi_i} - d_i, \beta) & j \leq N \\ g'_{\beta_{\xi_i-N}}(q_{\xi_i} - d_i, \beta) & N < j \leq N + k \\ 0 & N + k < j \end{cases} \]  

\[ \frac{\partial \mu(i, \theta)}{\partial \theta_j} = \frac{\partial g(q_{\xi_i} - d_i, \beta)}{\partial \theta_j} \]  

\[ \frac{\partial \tau(i, \theta)}{\partial \theta_j} = \begin{cases} 2\alpha \sigma_\theta^2 g(q_{\xi_i} - d_i, \beta) g(q_{\xi_i} - d_i, \beta)^{2\alpha-1} & j \leq N + k \\ 2\sigma_\theta g(q_{\xi_i} - d_i, \beta)^{2\alpha} & j = N + k + 1 \\ 2\sigma_\theta^2 g(q_{\xi_i} - d_i, \beta)^{2\alpha} \log g(q_{\xi_i} - d_i, \beta) & j = N + k + 2 \end{cases} \]
\[
J_{j,j} = \sum_{i \mid s_i = j} \frac{(g'_x(q_s_i - d_i, \beta))^2}{\sigma_0^2 g(q_s_i - d_i, \beta)^{2\alpha}} + \frac{1}{2} \sum_{i \mid s_i = j} \frac{4\alpha^2 (\sigma_0^2 g'_x(q_s_i - d_i, \beta) g(q_s_i - d_i, \beta)^{2\alpha-1})^2}{\sigma_0^4 g(q_s_i - d_i, \beta)^{4\alpha}}
\]

(15)

For the polynomial nonlinearity \( g_P(x, \beta) = \beta_1 x + \beta_2 x^2 + \ldots + \beta_k x^k \) considered here we have

\[
g'_P(x, \beta) = \beta_1 + 2\beta_2 x + \ldots + k\beta_k x^{k-1}
\]

\[
g'_{\beta_j}(x, \beta) = x^j.
\]

For the sigmoidal nonlinearity \( g_S(x, \beta) = \beta_1 + \frac{\beta_2}{1 + \exp(-\beta_3 x)} \) we have

\[
g'_S(x, \beta) = \frac{\beta_2 \beta_3 \exp(-\beta_3 x)}{(1 + \exp(-\beta_3 x))^2}
\]

\[
g'_{\beta_1}(x, \beta) = 1
\]

\[
g'_S(x, \beta) = \frac{1}{1 + \exp(-\beta_3 x)}
\]

\[
g'_{\beta_3}(x, \beta) = \frac{\beta_2 x \exp(-\beta_3 x)}{(1 + \exp(-\beta_3 x))^2}.
\]

(16)

REFERENCES