A Case Study on Monitoring Polynomial Profiles in the Automotive Industry

Amirhossein Amiri, Willis A. Jensen and Reza Baradaran Kazemzadeh

In some statistical process control applications, the quality of a process or product can be characterized by a relationship between a response variable and one or more explanatory variables, which is referred to as profile. We consider here profiles obtained from the automotive industry. One of the most important quality characteristics of an automobile engine is the relationship between the torque produced by an engine and the engine speed in revolutions per minute (RPM). For this particular engine type (TU3, assembled in the Peugeot automobile), the engine is run at different RPM values and the corresponding torque values are obtained. In other words, the torque produced by each engine is considered as a response variable and the correspondent speed values are considered as an explanatory variable. If the manufacturing process is in control, the profiles that describe the relationship between RPM and torque should be similar. An engine with mechanical defects or other issues will yield a profile that is different from the good engines. It is desirable to implement a procedure to monitor the quality level of the engines. Because there are multiple RPM values obtained for each engine it is natural to try to implement a multivariate quality control procedure to detect engines that are not acceptable. However, if the number of RPM values is large, it will be better to fit a parametric model to describe the relationship and monitor the estimated parameters in place of the actual torque values.

In Phase I of a control chart scheme, we have a historical data set at our disposal from which we are trying to determine if the profiles form a stable set of similar profiles with no outlying profiles. Once we are satisfied that the data set contains no outlying profiles, we can estimate the process parameters that are then fixed for Phase II of the control chart scheme. In Phase II, we are monitoring the process and are interested in detecting shifts as quickly as possible for real-time profiles that we continue to collect. Thus, our goal in Phase II is to know when the process has shifted from its historical, stable state so that we can quickly react to those changes and prevent problems.

A good introduction and literature review on profile monitoring can be found in Woodall et al. A more recent review can be found in Woodall. A wide variety of literature has appeared in recent years indicating a growing popularity in the approach. Several applications of profiles monitoring have been reported by authors such as Kang and Albin, Mahmoud and Woodall, Woodall et al., Wang and Tsung, Montgomery, Woodall, Zou et al., and Williams et al. Simple linear profiles for Phase I and

1. Introduction

Sometimes, quality of a process or product is characterized by a relationship between a response variable and one or more explanatory variables, which is referred to as profile. We consider here profiles obtained from the automotive industry. One of the most important quality characteristics of an automobile engine is the relationship between the torque produced by an engine and the engine speed in revolutions per minute (RPM). For this particular engine type (TU3, assembled in the Peugeot automobile), the engine is run at different RPM values and the corresponding torque values are obtained. In other words, the torque produced by each engine is considered as a response variable and the correspondent speed values are considered as an explanatory variable. If the manufacturing process is in control, the profiles that describe the relationship between RPM and torque should be similar. An engine with mechanical defects or other issues will yield a profile that is different from the good engines. It is desirable to implement a procedure to monitor the quality level of the engines. Because there are multiple RPM values obtained for each engine it is natural to try to implement a multivariate quality control procedure to detect engines that are not acceptable. However, if the number of RPM values is large, it will be better to fit a parametric model to describe the relationship and monitor the estimated parameters in place of the actual torque values.

In Phase I of a control chart scheme, we have a historical data set at our disposal from which we are trying to determine if the profiles form a stable set of similar profiles with no outlying profiles. Once we are satisfied that the data set contains no outlying profiles, we can estimate the process parameters that are then fixed for Phase II of the control chart scheme. In Phase II, we are monitoring the process and are interested in detecting shifts as quickly as possible for real-time profiles that we continue to collect. Thus, our goal in Phase II is to know when the process has shifted from its historical, stable state so that we can quickly react to those changes and prevent problems.

A good introduction and literature review on profile monitoring can be found in Woodall et al. A more recent review can be found in Woodall. A wide variety of literature has appeared in recent years indicating a growing popularity in the approach. Several applications of profiles monitoring have been reported by authors such as Kang and Albin, Mahmoud and Woodall, Woodall et al., Wang and Tsung, Montgomery, Woodall, Zou et al., and Williams et al. Simple linear profiles for Phase I and
Phase II applications were addressed by Mestek et al., Stover and Brill, Kang and Albin, Kim et al., Mahmoud and Woodall, Noorossana et al., Wang and Tsung, Gupta et al., Zou et al., Mahmoud et al., Zhang et al., and Saghaei et al.

There have been several recent papers dealing with extensions to profiles that can be modeled by a general linear model. Kazemzadeh et al. transformed a polynomial regression model to an orthogonal polynomial regression model and proposed a method based on using EWMA control charts to monitor the parameters of orthogonal polynomial regression model in Phase II. Kazemzadeh et al. proposed Phase I methods for monitoring a kth-order polynomial profiles. They also compared the performance of their methods under both step and isolated shifts in the regression parameters. Mahmoud proposed a Phase I methods for multiple linear profiles and compared the F-statistic method of Mahmoud and Woodall with a T\(^2\)-based control chart in Phase I.

In all the previous references, it is assumed that the error terms in the linear model are independently and identically distributed normal random variables. However, in certain cases these assumptions can be violated. Noorossana et al. investigated the effect of non-normality of the error terms on the performances of the EWMA/R method by Kang and Albin. Jensen et al. proposed a linear mixed model (LMM) to account for the autocorrelation within a linear profile in Phase I. Jensen and Birch showed that the use of mixed models could have significant advantages when there is autocorrelation within nonlinear regression models. Soleimani et al. proposed a transformation to eliminate the AR(1) structure between observations within a simple linear profiles and then used the traditional control procedures in the literature to monitor the profiles in Phase II. Noorossana et al. and Kazemzadeh et al. considered the linear and polynomial profiles over time and modeled autocorrelation between profiles as a first-order autoregressive process, respectively. In other words, it is assumed that the time is short between profiles, thus the corresponding observations in two successive profiles are autocorrelated. They proposed some methods based on a time-series approach and evaluated the performance of their methods.

In this case study we focus on a Phase I application of profile monitoring and show the step-by-step analysis we performed on the data. The structure of the paper is as follows: In the next section, it is shown that the relationship between the torque and the RPM of an engine can be modeled by a second-order polynomial profile. In addition, we discuss that there is an autocorrelation structure between residuals within a profile and as a result using a standard approach is inefficient. An alternative approach for modeling this relationship that can take into account the autocorrelation within a profile is discussed in Section 3. In Section 4, using the alternative approach, a Phase I study is conducted to check the stability of the process and determine if there are any outlying profiles. Then, the mean vector and variance–covariance matrix of regression parameters are estimated by using the observations from a stable process. Our concluding remarks are given in the final section.

2. Model and structure of the process

In the literature of profile monitoring in Phase I, it is assumed that there are m samples of size n\(_j\) in the form of \((x_{ij}, y_{ij}), i = 1, 2, \ldots, n_j, j = 1, 2, \ldots, m\). In other words, the subscript i represents the ith observation within each profile and subscript j shows the jth profile collected over time. One approach in the literature is referred to as a parametric approach, where one estimates the regression parameters after fitting an adequate parametric model and then uses the vector of regression parameters in a suitable statistic. This approach of fitting parametric regression models to the profiles is used throughout the paper.

For the engine data, we have a total of 26 engines at our disposal for Phase I analysis. In each engine, the speed values are set equal to 1500, 2000, 2500, 2660, 2800, 2940, 3500, 4000, 4500, 5000, 5225, 5500, 5775, and 6000 RPM and its corresponding torque values are measured. Hence, 14 points are gathered for each engine and form the profile of interest. The data set for 26 engines was collected and is shown in Appendix A. As with many situations, it is important to ensure that the fitted model is appropriate for the data at hand. A good modeler will try a variety of potential models, trying to keep the models as simple as possible while still explaining a significant percentage of the variability in the data. In our situation, a quadratic polynomial turns out to relatively simple model that explains the data well. That is not to say that an alternative model could not be fitted which would explain more of the variability of the data. Thus, for each profile we fit the following model as:

\[ y_j = \beta_0 + \beta_1 x_j + \beta_2 x_j^2 + e_j \]  

(1)

where \(y_j\) is the vector containing the torque values, \(x_j\) is the vector containing the RPM values, and \(e_j\) is the vector containing the errors. Because the RPM values are the same for all the profiles we have \(x_j = x\) for all values of \(j\). The parameters \(\beta_0\), \(\beta_1\), and \(\beta_2\) are estimated for each profile. After fitting this model, we obtain high values of the adjusted coefficient of determination for all engines. The scatterplot for one engine (Engine number 329) is illustrated in Figure 1 as an example. As is shown in Figure 1, the speed values are mean corrected to decrease the effect of multicollinearity. In other words, we replace \(x_j\) with \(x_j - \bar{x}\) in the above equation. The adjusted coefficient of determination \(R^2\) (adjusted) is equal to 96.3% and the model fits well. Similar results (not reported here) were obtained for the other 25 investigated engines.

The estimates of the regression parameters, the variance inflation factors (VIF), and analysis of variance table for engine number 329 are shown in Table I.

Based on the analysis of variance as well as separate individual tests in Table I, the significance of the regression parameters is confirmed. The VIFs show that there is no multicollinearity between explanatory variables. The other engines were evaluated using the above procedure and the adjusted coefficients of determination and estimates of regression parameters are summarized in Table II.
Figure 1. A second-order polynomial profile fit for engine number 329. This figure is available in colour online at www.interscience.wiley.com/journal/qre

Table I. Regression analysis results for Engine 329 for the model fit from Equation (1) with torque as the response and RPM as the explanatory variable

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>T</th>
<th>P</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>112.992</td>
<td>104.02</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>RPM</td>
<td>-0.0061551</td>
<td>-13.11</td>
<td>0.00</td>
<td>1.0</td>
</tr>
<tr>
<td>RPM²</td>
<td>-0.00000522</td>
<td>-12.67</td>
<td>0.00</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Analysis of variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>2</td>
<td>2156.2</td>
<td>1078.1</td>
<td>169.12</td>
<td>0.00</td>
</tr>
<tr>
<td>Residual</td>
<td>11</td>
<td>70.1</td>
<td>6.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>13</td>
<td>2226.3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The independence of residuals over time for each engine is one of the assumptions of regression analysis when one uses ordinary least square (OLS) method. We used a run chart to check this assumption. Run charts for standardized residuals of four engines are depicted in Figure 2. The results show that the clustering and trend hypothesis tests are significant and as a result it can be concluded that the residuals are correlated. Similar results are obtained for the other engines (the results are not reported here).

Model adequacy checking showed that we are faced with a case in which there is autocorrelation between residuals and as a result between observations in each profile. The least-square method is not recommended to estimate the regression parameters in this case because it ignores the presence of autocorrelation. As a result we propose an alternative modeling approach to the standard approach to address the limitations of the OLS approach.

3. An alternative approach

The proposed approach that we consider here will account for the autocorrelation within profiles. Note that this approach is simply a method for reducing the dimensionality of the data set prior to using a control chart to determine outlying profiles. The approach is an LMM approach, as was proposed by Jensen et al.\textsuperscript{22}. A good introduction to the LMM can be found in Verbeke and Molenberghs\textsuperscript{27}. An LMM treats the profiles as a random sample of profiles and allows for the inclusion of autocorrelation within a profile. The LMM is given as

$$y_j = X\beta + Zb_j + e_j,$$

where $y_j$ is the response vector for the $j$th profile, $X$ is the observation matrix, $\beta$ is the vector of fixed parameters, $Z$ is the matrix associated with the random effects, and $b_j$ is the vector of the random effects, which needs to be obtained from the data. The random effects indicate the deviation in the fixed parameters ($\beta$) for a particular profile. We assume that the random effects are multivariate normally distributed, that is $b_j \sim \text{MN}(0,D)$, where $0$ is a zero vector and $D$ is a positive-definite variance–covariance matrix with all off-diagonal elements equal to 0. In other words, the random effects are uncorrelated with each other. To allow for the correlated errors, we assume that $e_j \sim \text{MN}(0,\Sigma)$, where $\Sigma$ is a
positive-definite matrix. In addition, we assume that the random effects and errors are independent of each other, that is Cov(b_j, e_j) = 0. It is common to allow Σ to have some structure (AR(1) and compound symmetry for example) that reduces the number of parameters that need to be estimated in the variance–covariance matrix. The choice of which structure to use will depend on the data and we will use Draftman’s display as discussed by Jensen et al.\(^{22}\) to obtain the appropriate structure for this data set.

To implement the LMM approach on the 26 profiles, we first determine which of the regression parameters need a random effect. From the estimated fixed effects shown in Table II from the OLS approach, we computed the sample standard deviation for the intercept, linear, and quadratic terms. To determine the autocorrelation structure within the profile, we use Draftman’s display as discussed in Jensen et al.\(^{22}\) to obtain the appropriate structure for this data set.

The large standard deviation for the intercept indicates that just one random effect may be needed. However, initial attempts to fit an LMM on these data with just one random effect for the intercept were unsuccessful because of lack of convergence. This is because the intercept estimates are many orders of magnitude larger than the other terms. Thus, we had to scale the RPM values so that the estimates are of similar magnitude, thereby ensuring convergence of the LMM method. This is done by dividing the adjusted (by subtracting the mean) RPM values by 10,000. With this adjustment to ensure convergence, we are able to include random effects for the intercept, slope, and quadratic terms. To determine the autocorrelation structure within the profile, we use Draftman’s display as discussed in Jensen et al.\(^{22}\) This display is a plot of the within-profile residuals (centered and scaled by torque) at RPM\(_j\) versus those at RPM\(_{j+1}\) for all levels of RPM. For this engine data, Draftman’s display is shown in Figure 3. We see that the plots at the diagonal show a positive correlation between successive RPM levels. As we move away from the diagonal, the correlation decreases indicating that the RPM levels further apart are not as strongly correlated. Thus, an AR(1) structure for the \(\Sigma\) matrix is appropriate to model the autocorrelation of the errors within a profile. Note that the residuals used here are based on \((y_j - \hat{\beta}^T x - Z\hat{b})\), computed under the assumption of uncorrelated errors, which means that the residuals represent the within-profile variability.

AS the variance–covariance matrices \(D\) and \(\Sigma\) are not known, we obtain the estimated variance–covariance matrices by restricted maximum likelihood estimation using PROC MIXED in SAS. To do so we assume that the AR(1) structure and then obtain the best estimates given that assumption. Given an estimate of \(D\) and \(\Sigma\), we can compute, \(\hat{\beta}\), the vector of fixed effects as

\[
\hat{\beta} = (X^T \hat{\Sigma}^{-1} X)^{-1} X^T \hat{\Sigma}^{-1} \hat{\mu}
\]

where \(\hat{\Sigma} = Z \hat{D} Z' + \hat{\Sigma}\) is the overall estimated variance–covariance matrix and \(\hat{\mu}\) is the average of the response vectors. The estimated random effects (also known as the eblups because they are estimated as best linear unbiased predictors) are then given by

\[
\hat{b}_j = \hat{D} Z' \hat{\Sigma}^{-1} (y_j - X \hat{\beta})
\]
Thus, fitting the LMM to our data results in a vector containing the estimates of fixed effects, $\hat{\beta}$, that is common for all profiles as well as a vector containing the random effects, $\hat{\beta}_j$, which is unique for each profile.

For the control chart procedure, Jensen et al.\textsuperscript{22} showed that only the eblups ($\hat{\beta}_j$) are needed to determine the stability of the process. We obtained the eblups for the 26 engine profiles, which are given in Table IV. A profile plot showing the fitted profiles based on the eblups is shown in Figure 4.

As a check that the fitted LMM is appropriate, we show in Figure 5, a histogram of the residuals as an exploratory measure to determine if normality is a reasonable assumption. The residuals for the 26 profiles combined are shown here and are based on $(y_j - X\hat{\beta} - Z\hat{\beta}_j)$ as was done in Draftman’s display from Figure 3, with the exception that we do not assume uncorrelated errors. Rather, we use the estimated variance–covariance matrix using the AR(1) structure to obtain the residuals. We do not do a formal test of normality which would be invalid for these data because the residuals are not independent of each other but rather show this histogram as a check that the normality assumption appears to be reasonable. While there is a slight amount of left skewness in the data, the residuals are centered around zero and we believe that the normality assumption is adequate.

4. Phase I studies

From the previous section, we have now shown a way that the profiles can be reduced to a smaller set of parameter estimates, which can be analyzed to determine if the Phase I data are in control. We will use a $T^2$-based control procedure to investigate
the stability of the process as well as determine if there are any outlying profiles of the 26 profiles. Once we have removed any outlying profiles, we will be able to estimate the mean vector and covariance matrix of the process to be used to obtain control limits for Phase II of the control chart scheme.

As noted by Vargas\textsuperscript{28} and Jensen \textit{et al.}\textsuperscript{29}, there are multiple versions of the $T^2$ control chart that can be used for Phase I studies. Based on the recommendation of Figure 3 of Jensen \textit{et al.}\textsuperscript{29}, we will use a robust $T^2$ control chart based on the minimum volume ellipsoid (MVE) as well as one based on the successive differences (SD) as shown in Vargas\textsuperscript{28}. The MVE control chart is good for detecting outliers as well as small clusters of outliers. However, the SD control chart is good for detecting step changes in the profiles. The use of both of these two charts together will allow us to signal a wider variety of signals in the Phase I data.

The general form of $T^2$ statistic for the SD chart for profile $j$ is

$$T^2_j = (\hat{y}_j - \tilde{y})' S_{SD}^{-1} (\hat{y}_j - \tilde{y})$$

where $S_{SD}$ is calculated as follows:

$$S_{SD} = \frac{1}{2(m-1)} \sum_{j=1}^{m-1} (\hat{y}_{j+1} - \hat{y}_j) (\hat{y}_{j+1} - \hat{y}_j)'$$

In these equations, $\hat{y}_j$ represents the estimated vector containing the model parameters from the LMM approach and $\tilde{y}$ represents the mean of the estimated vectors for all the profiles. For the LMM approach, we replace the $\hat{y}_j$ in Equations (4) and (5) with the estimated vector of random effects, $\hat{b}_j$, from Equation (3). As discussed by Jensen \textit{et al.}\textsuperscript{22}, the average of the
Table IV. The eblups from the LMM analysis of the 26 engine profiles

<table>
<thead>
<tr>
<th>Engine No.</th>
<th>( \hat{b}_0 )</th>
<th>( \hat{b}_1 )</th>
<th>( \hat{b}_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>329</td>
<td>0.953</td>
<td>-0.532</td>
<td>0.253</td>
</tr>
<tr>
<td>449</td>
<td>-0.381</td>
<td>0.297</td>
<td>-0.155</td>
</tr>
<tr>
<td>529</td>
<td>0.986</td>
<td>-0.019</td>
<td>0.047</td>
</tr>
<tr>
<td>642</td>
<td>-1.19</td>
<td>0.769</td>
<td>-0.36</td>
</tr>
<tr>
<td>724</td>
<td>-1.936</td>
<td>0.874</td>
<td>-0.421</td>
</tr>
<tr>
<td>803</td>
<td>-0.614</td>
<td>-0.093</td>
<td>0.01</td>
</tr>
<tr>
<td>930</td>
<td>0.373</td>
<td>-0.241</td>
<td>0.143</td>
</tr>
<tr>
<td>1148</td>
<td>0.102</td>
<td>-0.097</td>
<td>0.009</td>
</tr>
<tr>
<td>1171</td>
<td>-0.28</td>
<td>0.245</td>
<td>-0.129</td>
</tr>
<tr>
<td>1516</td>
<td>0.387</td>
<td>-0.574</td>
<td>0.281</td>
</tr>
<tr>
<td>1791</td>
<td>0.715</td>
<td>-0.926</td>
<td>0.458</td>
</tr>
<tr>
<td>2600</td>
<td>1.012</td>
<td>-0.502</td>
<td>0.233</td>
</tr>
<tr>
<td>3100</td>
<td>0.37</td>
<td>-0.265</td>
<td>0.161</td>
</tr>
<tr>
<td>3720</td>
<td>-0.412</td>
<td>0.65</td>
<td>-0.285</td>
</tr>
<tr>
<td>4025</td>
<td>-0.378</td>
<td>-0.054</td>
<td>0.034</td>
</tr>
<tr>
<td>4068</td>
<td>0.471</td>
<td>-0.763</td>
<td>0.341</td>
</tr>
<tr>
<td>4926</td>
<td>1.998</td>
<td>-1.002</td>
<td>0.485</td>
</tr>
<tr>
<td>5155</td>
<td>1.099</td>
<td>-0.441</td>
<td>0.259</td>
</tr>
<tr>
<td>6143</td>
<td>-0.992</td>
<td>0.823</td>
<td>-0.411</td>
</tr>
<tr>
<td>6844</td>
<td>-0.419</td>
<td>0.342</td>
<td>-0.197</td>
</tr>
<tr>
<td>7811</td>
<td>-1.209</td>
<td>0.567</td>
<td>-0.296</td>
</tr>
<tr>
<td>8007</td>
<td>-0.715</td>
<td>0.271</td>
<td>-0.127</td>
</tr>
<tr>
<td>8623</td>
<td>0.102</td>
<td>0.118</td>
<td>-0.035</td>
</tr>
<tr>
<td>9388</td>
<td>1.146</td>
<td>-0.379</td>
<td>0.184</td>
</tr>
<tr>
<td>9404</td>
<td>-1.853</td>
<td>1.07</td>
<td>-0.565</td>
</tr>
<tr>
<td>10430</td>
<td>0.667</td>
<td>-0.138</td>
<td>0.079</td>
</tr>
</tbody>
</table>

Figure 4. Profile plot showing the fitted LMM model to all 26 profiles. This figure is available in colour online at www.interscience.wiley.com/journal/qre

Random effects vectors will be zero, thus we can rewrite Equation (4) as follows:

\[
T_j^2 = (\hat{b}_j)' \left[ \frac{1}{2(m-1)} \sum_{j=1}^{m-1} (\hat{b}_{j+1} - \hat{b}_j)(\hat{b}_{j+1} - \hat{b}_j)' \right]^{-1} (\hat{b}_j) \tag{6}
\]
In multivariate control charts, if \( m > p^2 + 3p \) (\( m \) is number of samples and \( p \) is number of quality characteristics), the \( T^2 \) statistic based on successive differences follows a chi-square distribution with \( p \) degrees of freedom, asymptotically (Williams et al. 30). Hence, the upper control limit (UCL) for the statistic in Equation (6) is as follows:

\[
UCL = \chi^2_{p, \alpha}
\]  

(7)

As the underlying model is a second-order polynomial profile, the number of regression parameters is equal to three and \( m \) for the investigated case is equal to 26. Hence, \( m = 26 \) is greater than \( p^2 + 3p = 18 \) in our case and the UCL in Equation (7) can be used here for the \( T^2 \) control chart based on successive differences. The UCL for \( T^2 \) statistics based on the SD considering significance level of \( \alpha = 0.00197 \) and 3 degrees of freedom is calculated as follows:

\[
UCL = \chi^2_{3, 0.00197} = 14.83
\]  

(8)

For the \( T^2 \) chart based on the MVE, we use general form of \( T^2 \) statistic for the MVE chart for profile \( j \), which is

\[
T^2_j = (\hat{b}_j - \hat{b}_{MVE})' S^{-1}_{MVE} (\hat{b}_j - \hat{b}_{MVE})
\]  

(9)

where \( \hat{b}_{MVE} \) and \( S_{MVE} \) are the center and variance–covariance matrix of the ellipsoid obtained from the data. The UCL for this chart must be obtained by simulation.

Because we are using two different control charts, one based on SD and one based on MVE, we need to control the overall significance level of the procedure. To obtain an overall significance level of \( \alpha = 0.1 \), each control chart should have roughly the \( \alpha \) equal to 0.05. This probability is calculated as follows:

\[
\alpha = 1 - (1 - \alpha_{overall})^{1/2} = 1 - (1 - 0.1)^{1/2} \approx 0.05
\]  

(10)

Similarly, to obtain an overall significance level of \( \alpha = 0.05 \) for each control chart, each statistic should have the following probability of Type I error:

\[
\alpha = 1 - (1 - \alpha_{overall})^{1/m} = 1 - (1 - 0.05)^{1/26} = 0.00197
\]  

(11)

The UCL for \( T^2 \) control chart based on MVE for \( m \) equal to 26, \( p \) equal to 3, and \( \alpha = 0.05 \) was given as 31.76 in Jensen et al. 29, who obtained the value by simulation. The \( T^2 \) statistics based on the SD and MVE for LMM approach were computed and the corresponding \( T^2 \) control charts are shown in Figures 6 and 7.
The $T^2$ control charts based on both SD and MVE for the LMM approach show that the process is in statistical control and stable (Figures 6 and 7). If the process was not stable, the assignable cause should be detected and the data values related to that assignable cause should be omitted. It should be noted that if one statistic falls out of the UCL, one should first omit the corresponding observation, then recalculate the covariance matrix of observations and estimate the regression parameters for the remaining observations. This procedure is repeated until no assignable causes are detected.

As the process is stable, we can use the data set to estimate the parameters of the process including mean vector and variance–covariance matrix. These parameters can be used to construct the control charts for Phase II studies. The mean vector and variance–covariance matrix are estimated by using MVE estimator and the results are given as follows:

$$\hat{b}_{MVE} = \left(\begin{array}{c} 0 \\ 0 \end{array}\right), \quad S_{MVE} = \left(\begin{array}{ccc} 0.94771 & -0.49118 & -0.24681 \\ -0.49118 & 0.32360 & -0.15933 \\ -0.24681 & -0.15933 & 0.07902 \end{array}\right)$$

(12)

The estimated parameters in Equation (12) can be used in Phase II to monitor the process in the future. We denote $\hat{b}_{m+1}$ as the estimate of a future observation, which is obtained by replacing $y_j$ with $y_{m+1}$ in Equation (3). Note that we use the adjusted RPM values as well as the original 26 in-control observations in the estimation of the vector of fixed effects and variance-covariance matrices in Equation (3) because there were no out-of-control observations in either of the control charts. The following $T^2$-squared statistic can be applied in this phase as

$$T^2_{m+1} = (\hat{b}_{m+1} - \mu) \Sigma^{-1} (\hat{b}_{m+1} - \mu)$$

(13)

where $\mu$ and $\Sigma$ are replaced by $\hat{b}_{MVE}$ and $S_{MVE}$ in Equation (12), respectively. For each future observation, $\hat{b}_j$, which is the estimator vector of the random effects, is computed by using Equation (3). Then, $T^2$-squared statistic is computed by using Equation (13) and is plotted on the $T^2$ control chart. In Phase II, the $T^2$-squared statistic in Equation (13) follows a chi-squared distribution with $p$ degrees of freedom, which is equal to three in our case. Hence, to achieve the probability of Type I error equal to $\alpha$, the UCL for the statistic in Equation (13) is equal to $\chi^2_{p, \alpha} / p$. Using the mentioned $T^2$ control chart, one can monitor the process of producing the engine in the future and decide whether or not process is in statistical control.

5. Conclusions

In this paper, the relationship between the torque produced by an automobile engine and its speed was investigated. Our studies showed that this relationship can be modeled by a second-order polynomial profile and there is autocorrelation structure within each profile. We proposed LMM approach to estimate polynomial regression parameters. Then, we used a $T^2$-based control procedure to conduct Phase I studies. The results showed that the process that produces the engine is stable. Furthermore, using the stable process, the mean vector and variance–covariance matrix of parameter estimators, which are required for monitoring the process in the future, were estimated. Finally, we discussed how one can use these estimators to form a $T^2$ based control chart to monitor polynomial profiles in Phase II.

Acknowledgements

The authors are grateful to the management of Engine Shop 3 in Iran Khodro Corporation for providing the data set used in the paper. In addition, we appreciate very much the comments from the editor and an anonymous referee that helped us to correct some errors, clarify the main points, and improve this paper.
References


Appendix: A Historical Data Set of 26 Automobile Engines (Torque versus RPM)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1500</td>
<td>98.53</td>
<td>96.35</td>
<td>98.77</td>
<td>96.70</td>
<td>96.75</td>
<td>97.61</td>
<td>100.06</td>
<td>94.55</td>
<td>96.48</td>
<td>100.73</td>
</tr>
<tr>
<td>2000</td>
<td>102.65</td>
<td>100.74</td>
<td>103.03</td>
<td>100.05</td>
<td>100.87</td>
<td>102.46</td>
<td>103.60</td>
<td>103.22</td>
<td>100.87</td>
<td>103.82</td>
</tr>
<tr>
<td>2500</td>
<td>113.82</td>
<td>110.67</td>
<td>111.99</td>
<td>111.17</td>
<td>110.14</td>
<td>112.18</td>
<td>112.74</td>
<td>112.99</td>
<td>110.81</td>
<td>113.46</td>
</tr>
<tr>
<td>2660</td>
<td>115.26</td>
<td>113.06</td>
<td>112.78</td>
<td>111.51</td>
<td>110.48</td>
<td>112.99</td>
<td>113.56</td>
<td>114.18</td>
<td>113.20</td>
<td>113.69</td>
</tr>
<tr>
<td>2800</td>
<td>116.24</td>
<td>114.58</td>
<td>113.14</td>
<td>112.01</td>
<td>110.94</td>
<td>114.54</td>
<td>112.85</td>
<td>116.48</td>
<td>114.73</td>
<td>115.07</td>
</tr>
<tr>
<td>2940</td>
<td>117.06</td>
<td>114.98</td>
<td>113.73</td>
<td>111.23</td>
<td>111.17</td>
<td>115.00</td>
<td>114.49</td>
<td>115.33</td>
<td>115.13</td>
<td>115.32</td>
</tr>
<tr>
<td>3500</td>
<td>109.89</td>
<td>108.55</td>
<td>110.30</td>
<td>105.64</td>
<td>105.78</td>
<td>108.99</td>
<td>108.95</td>
<td>109.59</td>
<td>108.69</td>
<td>109.71</td>
</tr>
</tbody>
</table>
## Authors' biographies

**Amirhossein Amiri** is an Assistant Professor at Shahed University. He holds a BS, MS, and PhD in Industrial Engineering from Khajeh Nasir University of Technology, Iran University of Science and Technology, and Tarbiat Modares University, respectively. He is a member of the Iranian Statistical Association. His research interests are statistical quality control, profile monitoring, and Six Sigma.

**Willis A. Jensen** is an Associate in the Medical Products Division at W. L. Gore & Associates, where he provides statistical support and training across the division. He holds a BS and MS in Statistics from Brigham Young University and a PhD in Statistics from Virginia Tech. He is a member of the American Society for Quality and the American Statistical Association. His research interests include all aspects of industrial statistics, including statistical process control and profile monitoring and his publications have appeared in *Quality and Reliability Engineering International* and the *Journal of Quality Technology*. 

---

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>4000</td>
<td>109.65</td>
<td>107.41</td>
<td>109.35</td>
<td>106.02</td>
<td>103.37</td>
<td>107.95</td>
<td>108.24</td>
<td>108.47</td>
<td>107.55</td>
</tr>
<tr>
<td>4500</td>
<td>105.72</td>
<td>103.90</td>
<td>107.61</td>
<td>103.11</td>
<td>102.23</td>
<td>103.65</td>
<td>105.56</td>
<td>105.27</td>
<td>104.03</td>
</tr>
<tr>
<td>5000</td>
<td>99.74</td>
<td>97.99</td>
<td>100.64</td>
<td>97.40</td>
<td>96.06</td>
<td>96.94</td>
<td>98.92</td>
<td>97.90</td>
<td>98.12</td>
</tr>
<tr>
<td>5225</td>
<td>95.97</td>
<td>94.27</td>
<td>97.59</td>
<td>93.88</td>
<td>92.39</td>
<td>92.78</td>
<td>95.41</td>
<td>94.67</td>
<td>94.39</td>
</tr>
<tr>
<td>5500</td>
<td>89.47</td>
<td>88.45</td>
<td>91.68</td>
<td>88.17</td>
<td>86.54</td>
<td>86.41</td>
<td>89.19</td>
<td>88.23</td>
<td>88.56</td>
</tr>
<tr>
<td>5775</td>
<td>81.96</td>
<td>81.44</td>
<td>84.58</td>
<td>81.18</td>
<td>79.31</td>
<td>78.60</td>
<td>81.85</td>
<td>80.86</td>
<td>81.54</td>
</tr>
<tr>
<td>6000</td>
<td>74.90</td>
<td>75.00</td>
<td>77.48</td>
<td>75.03</td>
<td>73.13</td>
<td>71.97</td>
<td>75.09</td>
<td>73.93</td>
<td>75.09</td>
</tr>
</tbody>
</table>

*Authors' biographies*
Reza Baradaran Kazemzadeh is an Associate Professor in Tarbiat Modares University. He received a BS in Operations Research, and a MS in Industrial Engineering. He holds a PhD in Technology Management from the University of Bradford. His research interests are quality engineering and management, profile monitoring, applied multivariate statistics, and operation management. His publications have appeared in Quality & Quantity, European Journal of Operational Research, Communications in Statistics, and the International Journal of Advanced Manufacturing Technology.