AN EXAMPLE OF THE USE OF ANDREWS' PLOTS TO DETECT TIME VARIATION ETC (U)

JUL 81 A M HERZBERG

UNCLASSIFIED MRC-TSR-2239
AN EXAMPLE OF THE USE OF ANDREWS' PLOTS TO DETECT TIME VARIATIONS IN MODEL PARAMETERS AND OUTLYING OBSERVATIONS

Agnes M. Herzberg

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

July 1981

(Received June 23, 1981)

Approved for public release
Distribution unlimited

Sponsored by
U.S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709
AN EXAMPLE OF THE USE OF ANDREWS' PLOTS TO DETECT TIME VARIATIONS IN MODEL PARAMETERS AND OUTLYING OBSERVATIONS

Agnes M. Herzberg

Technical Summary Report #2239
July 1981

ABSTRACT

Andrews (1972) introduced a method of plotting high-dimensional data in two dimensions. This method is exploited as a graphical tool for the examination of changes over time in the parameters of a time series model. An example using a Fourier series model is given to illustrate the method. It is also shown how outlying observations in the data can be found.

AMS (MOS) Subject Classifications: 62M10, 62H30

Key Words: Andrews' plots, time series, outliers, spurious observations, exploratory analysis.

Work Unit Number 4 (Statistics and Probability)

†Imperial College of Science and Technology, University of London, London, U. K.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.
SIGNIFICANCE AND EXPLANATION

Andrews (1972) introduced a method of plotting high-dimensional data in two dimensions. In his method, Andrews represents each multidimensional point by a Fourier function. The clustering of plots of these functions is equivalent to the clustering of the multidimensional points. Andrews' method is exploited as a graphical tool for exploratory data analysis for the examination of changes over time in the parameters of a time series model. An example using the total Canadian unemployment figures from 1956-1975 is used to illustrate the method. These data have four spurious (outlying) observations and it is shown how these may be detected by the use of Andrews' plots.

The responsibility for the wording and views expressed in this descriptive summary lies with NRC, and not with the author of this report.
AN EXAMPLE OF THE USE OF ANDREWS' PLOTS TO DETECT TIME VARIATIONS IN MODEL PARAMETERS AND OUTLYING OBSERVATIONS

Agnes M. Herzberg†

1. Introduction

A graphical method is given for the examination of changes over time in the parameters of a time series model. This method can be used as an aid in exploratory data analysis. In a previous paper, Herzberg and Hickie (1981), the method is presented and two examples are given. A brief description of various multivariate graphical clustering methods and the use of Andrews' plots as a graphical tool in time series analysis is also given in Herzberg (1981). Here another model is used with a different set of data and further discussion given of the detection of outliers, or spurious observations.

2. Andrews' Plots

Andrews (1972) proposed the following simple and useful method of plotting high-dimensional data in two dimensions. If the data are m-dimensional, each point \( \mathbf{x} = (x_1, \ldots, x_m) \), where \( x_i \) (i = 1, ..., m) are the measured variables, is represented by the function

\[
f_x(t) = x_1 \cdot 2^{-1/2} + x_2 \sin t + x_3 \cos t + x_4 \sin 2t + x_5 \cos 2t + \ldots \tag{1}
\]

plotted over the range \(-\pi < t < \pi\). The functions given by (1) have several properties including the preservation of means, distances and variances and will also give one-dimensional projections. Thus, when (1) is plotted for

†Imperial College of Science and Technology, University of London, London, U. K.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.
each data point \( x \), the clustering of the points may be seen by a banding together of the plots of the functions. Tests of significance may also be made; see Herzberg and Hickie (1981).

3. Variation of Model Parameters

Herzberg and Hickie (1981) considered the regression model

\[
X_j = \xi \theta_j + \psi_j \quad (j = 1, \ldots, T-n+1),
\]

where \( T \) is the total number of observations, \( n \) is the number of observations in each subgroup of observations used for estimating the unknown parameters, \( X_j = (y_{1j}, \ldots, y_{nj})' \) is a \( n \times 1 \) vector, \( y_{ij} \) being the \( i \)th observation in the \( j \)th subgroup \( (i = 1, \ldots, n) \), \( \xi \) is the \( n \times m \) matrix of the regressors, \( \theta_j \) is the \( m \times 1 \) vector of unknown parameters to be estimated by least squares and \( \psi_j \) is the \( n \times 1 \) vector of error terms. All the elements of the \( \psi_j \)'s are assumed to be independent and normally distributed with mean \( 0 \) and variance \( \sigma^2 \). It is assumed that the \( T \) observations are taken sequentially over time and it is desired to examine the variation in the \( \theta_j \) over time.

Let \( \hat{\theta}_j = (\hat{\theta}_{1j}, \ldots, \hat{\theta}_{mj})' \) be the \( m \times 1 \) vector of least squares estimates of the elements of \( \theta_j \) obtained from the \( j \)th set of \( n \) observations \( (n < T) \), i.e. \( \hat{\theta}_1 \) is estimated from the first \( n \) observations, \( \hat{\theta}_2 \) is estimated from the second observation to the \( (n + 1) \)st observation, etc. From each \( \hat{\theta}_j \), a plot of the function \( f_{\hat{\theta}_j}(t) \), defined in (1), over the range \(-\pi < t < \pi\) was made. The plots of these functions will show the change over time in the vector of coefficients \( \theta_j \). The plots, \( f_{\hat{\theta}_j}(t) \), can be considered as a graphical weighted moving average. For each \( t \) a different weighting is given to the observations.
4. An Example

Table 1 shows the total Canadian unemployment figures from January 1956 to December 1975. It can be seen that the values for January 1958, 1961, 1971 and 1975 could be considered as being outliers or spurious observations in the data. Figure 1 gives a plot of these data.

The model

$$E(y_{j+i-1}) = \beta_{1j} + \beta_{2j}\sin\frac{2\pi i}{12} + \beta_{3j}\cos\frac{2\pi i}{12} + \beta_{4j}\sin\frac{4\pi i}{12} + \beta_{5j}\cos\frac{4\pi i}{12}$$

(i = 1, ..., 12; j = 1, ..., 229),

where $y_{j+i-1}$ is the observed unemployment figure in month $j+i-1$, was fitted to the data by least squares for each $j$ fixed and

$$\hat{\beta}_j = (\hat{\beta}_{1j}, \hat{\beta}_{2j}, \hat{\beta}_{3j}, \hat{\beta}_{4j}, \hat{\beta}_{5j})',$$

the least squares estimate of $\beta_j$ obtained.

The plots of the function

$$f_{\hat{\beta}_j}(t) = \hat{\beta}_{1j} + \hat{\beta}_{2j}\cos t + \hat{\beta}_{3j}\sin t + \hat{\beta}_{4j}\cos 2t + \hat{\beta}_{5j}\sin 2t$$

(j = 1, ..., 229),

were obtained and plotted. Note that (3) differs from (1) but the mathematical properties of (1) are retained. Several variations of (1) were tried but the outlying plots were most easily seen when (3) was used. This is due to the particular weighting which (3) gives to the $\hat{\beta}_{ij}$'s and thus to the individual observations.

It could be seen from the plots when plotted in chronological order on a graphics terminal that certain ones stood out from the others. Any long term increases or decreases in the plots were also noted.

The 229 Andrews' plots are given in Fig. 2 and Fig. 3. The plots in Fig. 2.k (k = 1, ..., 12) are those obtained from (3) for $j = k, k + 12, k + 24, ..., k + 108$. The plots in Fig. 2.k are similar except for the ones
Table 1. Total Canadian unemployment figures for 1956-1975 in 1000's.

<table>
<thead>
<tr>
<th>January</th>
<th>February</th>
<th>March</th>
<th>April</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>August</th>
<th>September</th>
<th>October</th>
<th>November</th>
<th>December</th>
</tr>
</thead>
<tbody>
<tr>
<td>1956</td>
<td>315</td>
<td>340</td>
<td>321</td>
<td>273</td>
<td>175</td>
<td>127</td>
<td>112</td>
<td>116</td>
<td>110</td>
<td>149</td>
<td>211</td>
</tr>
<tr>
<td>1957</td>
<td>328</td>
<td>352</td>
<td>378</td>
<td>334</td>
<td>209</td>
<td>177</td>
<td>181</td>
<td>194</td>
<td>214</td>
<td>223</td>
<td>318</td>
</tr>
<tr>
<td>1958</td>
<td>579</td>
<td>600</td>
<td>638</td>
<td>553</td>
<td>388</td>
<td>339</td>
<td>310</td>
<td>317</td>
<td>328</td>
<td>378</td>
<td>466</td>
</tr>
<tr>
<td>1959</td>
<td>577</td>
<td>570</td>
<td>553</td>
<td>466</td>
<td>354</td>
<td>248</td>
<td>239</td>
<td>257</td>
<td>224</td>
<td>250</td>
<td>316</td>
</tr>
<tr>
<td>1960</td>
<td>546</td>
<td>597</td>
<td>607</td>
<td>550</td>
<td>417</td>
<td>313</td>
<td>328</td>
<td>350</td>
<td>325</td>
<td>366</td>
<td>426</td>
</tr>
<tr>
<td>1961</td>
<td>690</td>
<td>716</td>
<td>702</td>
<td>619</td>
<td>454</td>
<td>367</td>
<td>351</td>
<td>320</td>
<td>305</td>
<td>315</td>
<td>347</td>
</tr>
<tr>
<td>1962</td>
<td>543</td>
<td>582</td>
<td>559</td>
<td>484</td>
<td>335</td>
<td>300</td>
<td>307</td>
<td>279</td>
<td>259</td>
<td>282</td>
<td>342</td>
</tr>
<tr>
<td>1963</td>
<td>541</td>
<td>546</td>
<td>550</td>
<td>463</td>
<td>347</td>
<td>305</td>
<td>294</td>
<td>271</td>
<td>251</td>
<td>266</td>
<td>303</td>
</tr>
<tr>
<td>1964</td>
<td>466</td>
<td>467</td>
<td>456</td>
<td>403</td>
<td>293</td>
<td>282</td>
<td>265</td>
<td>246</td>
<td>217</td>
<td>257</td>
<td>284</td>
</tr>
<tr>
<td>1965</td>
<td>407</td>
<td>397</td>
<td>387</td>
<td>371</td>
<td>265</td>
<td>257</td>
<td>244</td>
<td>211</td>
<td>176</td>
<td>171</td>
<td>220</td>
</tr>
<tr>
<td>1966</td>
<td>359</td>
<td>356</td>
<td>341</td>
<td>298</td>
<td>247</td>
<td>230</td>
<td>244</td>
<td>228</td>
<td>205</td>
<td>195</td>
<td>238</td>
</tr>
<tr>
<td>1967</td>
<td>381</td>
<td>396</td>
<td>400</td>
<td>365</td>
<td>304</td>
<td>292</td>
<td>284</td>
<td>247</td>
<td>219</td>
<td>254</td>
<td>289</td>
</tr>
<tr>
<td>1968</td>
<td>464</td>
<td>482</td>
<td>488</td>
<td>436</td>
<td>366</td>
<td>395</td>
<td>371</td>
<td>319</td>
<td>262</td>
<td>288</td>
<td>338</td>
</tr>
<tr>
<td>1969</td>
<td>467</td>
<td>473</td>
<td>448</td>
<td>432</td>
<td>386</td>
<td>383</td>
<td>349</td>
<td>318</td>
<td>279</td>
<td>314</td>
<td>354</td>
</tr>
<tr>
<td>1970</td>
<td>485</td>
<td>526</td>
<td>542</td>
<td>544</td>
<td>513</td>
<td>529</td>
<td>518</td>
<td>548</td>
<td>498</td>
<td>419</td>
<td>476</td>
</tr>
<tr>
<td>1971</td>
<td>668</td>
<td>675</td>
<td>650</td>
<td>659</td>
<td>543</td>
<td>551</td>
<td>514</td>
<td>455</td>
<td>434</td>
<td>447</td>
<td>503</td>
</tr>
<tr>
<td>1972</td>
<td>665</td>
<td>627</td>
<td>642</td>
<td>592</td>
<td>552</td>
<td>568</td>
<td>543</td>
<td>503</td>
<td>459</td>
<td>483</td>
<td>524</td>
</tr>
<tr>
<td>1973</td>
<td>688</td>
<td>655</td>
<td>608</td>
<td>570</td>
<td>493</td>
<td>503</td>
<td>461</td>
<td>433</td>
<td>421</td>
<td>429</td>
<td>468</td>
</tr>
<tr>
<td>1974</td>
<td>637</td>
<td>635</td>
<td>599</td>
<td>568</td>
<td>524</td>
<td>469</td>
<td>465</td>
<td>447</td>
<td>431</td>
<td>430</td>
<td>493</td>
</tr>
<tr>
<td>1975</td>
<td>817</td>
<td>839</td>
<td>840</td>
<td>795</td>
<td>714</td>
<td>704</td>
<td>653</td>
<td>623</td>
<td>586</td>
<td>576</td>
<td>640</td>
</tr>
</tbody>
</table>
FIG. 1. Canadian monthly total unemployment figures from 1956-1975 in 1000's.

PERSONS UNEMPLOYED X 10^-3

YEAR

Fig. 2. Andrews' plots, \( f_{eta_j}(t) \) (\( j = 1, \ldots, 120 \)) given by (3), obtained from (2) and sorted.

The darker curves denote these plots obtained in part from January 1958 or 1961.
FIG. 3. Andrews' plots, $f_{\widehat{B}}(t)$ ($j = 121, \ldots, 229$) given by (3), $\widehat{B}_j$ obtained from (2) and sorted.

Fig. 3. $k$ ($k = 1, \ldots, 12$) consists of plots for $j = k + 120$, $k + 132, \ldots, k + 228$ ($j \leq 229$).
(The darker curves denote these plots obtained in part from January 1971 or 1975.)
denoted by a thicker line. These are the ones whose coefficients are estimated from January 1958 or 1961. The plots in Fig. 3.k (k = 1, ..., 12) are those obtained from (3) for j = k + 120, k + 132, ..., k + 228 (j < 229). The plots in Fig. 3.k are similar except for the ones denoted by a thicker line. These are the ones whose coefficients are estimated from January 1971 or 1975.

Thus Andrews' plots can be used as a graphical method not only to examine changes over time in the parameters but also to detect abrupt changes in the observations reflected by changes in the parameters of the model over time. As mentioned elsewhere, the Andrews' plots can also be used to determine the period length when this is unknown.

5. Acknowledgement

This work was done while the author was a visitor at the Mathematics Research Center at the University of Wisconsin-Madison. The author would like to thank the good offices of the Mathematics Research Center and in particular Mr. F. W. Sauer who did the computer graphics.

Bibliography


Andrews' plots, time series, outliers, spurious observations, exploratory analysis

Andrews (1972) introduced a method of plotting high-dimensional data in two dimensions. This method is exploited as a graphical tool for the examination of changes over time in the parameters of a time series model. An example using a Fourier series model is given to illustrate the method. It is also shown how outlying observations in the data can be found.