Estimating the lump to fines composition split in iron ore production

J. E. Everett
Faculty of Economics and Commerce,
The University of Western Australia

Abstract

Iron ore is mined, crushed and separated into lump and fines components, which are sold as separate products. Quality (and therefore value) depends upon the shipped composition of each product matching a target composition vector, in iron and in several other minerals. Each day, ore is selected for mining to meet these target grades. The total composition of the candidate blocks for mining can be estimated from assays of material recovered from blast holes. But the lump and fines components will have different compositions, the lump usually being higher in iron and lower in other minerals. To estimate this lump to fines split, we need a “lump algorithm”, a vector predicting the lump percentage yield, and the lump to fines difference in each mineral. Ores of different geological types, or “geotypes” have systematically different lump algorithms.

Algorithms for specific ore types have been estimated by running the crusher with a single ore type over a sample period. This method is costly, interferes with production, and yields results of limited statistical power.

An alternative, described here, is to use Weighted-least-squares (WLS) multiple regression. This enables us to estimate an algorithm for each geotype, provided sample periods have sufficiently varied geotype mixes.

Keywords: quality control, decision support, mining, regression analysis, forecasting, weighted least squares.

1 Introduction

Iron ore is mined from open-cut pits and railed to the port, where it is crushed. The crushed ore is separated into lump and fines components, which are sold as separate products, for feed to blast furnace steel production.
The quality of the ore product depends upon it closely matching target composition, not only in iron, but also in phosphorus, silica and alumina. The customer blast furnaces are tuned to receive ore of the agreed target composition. Issues relating to quality control in iron ore production are discussed in [1,2].

When ore is crushed and split into lump and fines components, there are systematic differences between the lump and fines compositions. The lump component is usually (but not always) richer in iron, and with less silica and alumina, than the fines product. The lump and fines composition targets are set accordingly, to reflect the average difference between lump and fines compositions.

The proportion of lump generated, and the composition difference between lump and fines, is referred to as the “lump algorithm”. The composition of the original ore, before splitting into lump and fines components, is referred to as the “head grade”.

Before ore is mined, it is drilled for blasting. The drilled material is assayed and provides an estimate of the head grade, biased because the finer clay minerals tend to get lost in sampling the drilled material. The blasthole estimate is corrected for bias by a continuous exponentially smoothed comparison of the blasthole estimates and the corresponding port assays of crushed ore.

If the only objective was to produce head grade of target composition, then the areas of material for mining could be selected each shift to match the head grade target. However, we need to select ore that will split into lump and fines components, each matching the respective target composition.

It is known that ores of different geological types (different “geotypes”) have different lump algorithms. Some ore geotypes give a greater percentage of lump, or show a greater lump to fines composition difference than do other types of ore. Some geotypes even tend to split according to a “reverse algorithm”, with the lump component being lower in iron and higher in silica and alumina than is the fines component.

Accordingly, when candidate areas of ore for mining are being considered, we need to estimate not only the head grade, but also the lump algorithm.

Previously, algorithms for specific ore types have been estimated by running the crusher with a single ore type over a sample period. This method is costly. It is disruptive to the production flow, and therefore is infrequently carried out. It is a statistically unreliable method, because random variations within a geotype may be comparable to the systematic variations between geotypes, and the cost of interrupting normal production precludes the test being done with sufficient frequency to measure a representative range of ore.

An alternative approach, described here, is to consider the historical data for the mix of ore that has been actually crushed and sampled. The ore produced for each time period is routinely sampled after crushing, and the samples are assayed. Although each assay period contains ore from a mixture of geotypes, multiple regression analysis can ascribe lump algorithms to each geotype, provided that assay periods have varying proportions of geotypes (as is the case).

Interestingly, the approach is very closely equivalent to a study carried out by this author for the West Australian State Laundry Service [3].
government utility routinely washed linen for a number of government hospitals
and institutions. Each day’s wash included a variety of types of laundry, such as
nurses’ uniforms, bed sheets, operating theatre linen and so forth. Costs could be
ascribed to each shift, but not directly to each type of laundry, since they were
washed together. The problem was solved by a regression approach essentially
equivalent to the method being described here.

2 Multiple regression analysis

2.1 Ordinary least squares (OLS)

In normal production the crusher is fed with ore of a mixture of geotypes, and
periodic samples are taken of the lump and fines output.

Consider a set of geotypes \( j = 1 \ldots J \), where the \( j \) geotype “j” has mean
algorithm “G\(_j\)”. The algorithm G\(_j\) is a vector with five components, being the
lump percentage, and the lump-fines composition difference for iron,
phosphorus, silica and alumina.

Matching geotype input proportions with corresponding output lump and
fines product assays for successive assay periods \( n = 1 \ldots N \), we can formulate
the following linear equation (1), for the composite algorithm “A\(_n\)”, cumulated
across period \( n \). The residual \( e_n \) is the random variability, and the proportions
\( p_{jn} \) are the proportions of each geotype \( j \) in period \( n \).

\[
A_n = \sum_j p_{jn} G_j + e_n, \quad \text{where} \quad \sum_j p_{jn} = 1 \text{ for each } n \quad (1)
\]

\[
\sum e_n^2 = \sum_n (A_n - \sum_j p_{jn} G_j)^2 \quad (2)
\]

Provided the proportions \( p_{jn} \) vary across \( n \), standard multiple regression
techniques (with the constraint of zero intercept) can estimate values for \( G_j \), to
minimise the total \( \sum e_n^2 \) in equation (2).

2.2 Interpreting the regression

In the regression, the dependent variable is the vector \( A_n \), whose five components
are the proportion of lump tonnage produced in the period \( n \), and the difference
between the lump and fines assays for each of the four minerals, iron,
phosphorus, silica and alumina. There are thus, in effect, five separate
regressions, one for each of these vector components. The independent variables
for the regressions are the values \( p_{jn} \), the proportions of each geotype \( j \) crushed in
period \( n \). The coefficients returned by the regression analysis are the components
of the vectors \( G_j \), providing estimates of the lump algorithm for each geotype \( j \).

2.3 Allowing for variable error variance

The regression analysis, using equation (2), assumes \( \varepsilon^2 = E[e_n^2] \), the expected
variance of \( e_n \), is constant. This ordinary-least-squares (OLS) method gives
unbiased estimates of \( G_j \). However, if \( E[e_n^2] \) is not constant, but varies with \( n \),
the estimates are inefficient, and the uncertainty in the estimates is not minimised.

The tonnage crushed in each assay period varies, so we cannot assume the expected variance of \( e_n \) is constant. The problem of dealing with variable error variance is not fully discussed in most statistics textbooks, so will be briefly outlined here, for the general linear model. The model of equations (1) and (2) is an example of the general linear model.

Consider estimating the mean \( \mu_D \) of a quantity \( D \), given \( N \) multiple observations \( D_n \), each with an error \( d_n \), with variance \( \delta_n^2 \), differing between observations.

\[
D_n = \mu_D + d_n \tag{3}
\]

An unbiased, but inefficient estimator of \( \mu_D \) is:

\[
m_D' = \frac{\sum D_n}{N} = \frac{\sum (\mu_D + d_n)}{N} = \mu_D + \frac{\sum d_n}{N} \tag{4}
\]

A weighted mean, taking account of the individual variances, can provide a more efficient estimator of \( \mu_D \). Apply weights \( k_n \) to the \( n \)th observation, subject to the constraint:

\[
\sum k_n = 1 \tag{5}
\]

\[
m_D = \sum k_n D_n = \sum k_n \mu_D + \sum k_n d_n = \mu_D + \sum k_n d_n \tag{6}
\]

\[
\sigma^2[m_D] = E[\sum k_n d_n]^2 = \sum k_n^2 \delta_n^2 \tag{7}
\]

Using a Lagrange multiplier \( \lambda \), we can minimise the loss function (7) subject to the constraint (5):

\[
\frac{\partial[\sum k_n^2 \delta_n^2 - \lambda (\sum k_n - 1)]}{\partial w_j} = 2k_n \delta_n^2 - \lambda = 0 \tag{8}
\]

\[
k_n = \lambda / 2\delta_n^2 = \left(1/\delta_n^2\right)/\sum_m \left(1/\delta_m^2\right) \tag{9}
\]

Equation (9) shows that an efficient estimator requires estimates be used with a weight inversely proportional to their error variance.

Clearly, a sample from a larger production tonnage will have lower variance than a smaller tonnage sample. We will now show that, provided the errors have no serial correlation, the error variance is inversely proportional to the tonnage sampled. Consider compounding \( M \) observations \( D_m \) of equal variance \( \delta^2 \) (and negligible serial correlation) to form an average observation \( F \). \( F \) still has mean \( \mu_D \) but error \( f \), with variance \( \gamma \).

\[
F = \frac{\sum D_m}{M} = \mu_D + \frac{\sum d_m}{M} = \mu_D + f \tag{10}
\]
\[ f = \left( \sum d_m \right)/M \quad (11) \]

\[ \gamma^2 = E[f^2] = \left( \sum E[d_m^2]/M \right)/M^2 = \delta^2/M \quad (12) \]

So, assuming that the error variability has negligible serial correlation, the error variance is inversely proportional to the sample size. For the model of equation (1), the sample size is the production tonnage in each period.

### 2.4 Weighted least squares (WLS)

Following the above analysis, the Weighted Least Squares (WLS) regression method weights each observation in inverse proportion to its error variance. We have seen that, whereas OLS gives unbiased but inefficient estimates, WLS gives estimates that are both unbiased and efficient. WLS estimates (and OLS estimates) are unbiased because the expected error is zero. The WLS estimates are also efficient because their error variances are minimised.

The error variance is inversely proportional to the production tonnage. So individual records should be weighted, with weights inversely proportional to the variance, and therefore directly proportional to the production tonnage for each time interval. If \( W_n \) is the tonnage produced in period \( n \), we need to minimise \( \sum W_n e_n^2 \) in equation (14):

\[ \sum W_n e_n^2 = \sum n W_n (A_n - \sum_j p_{jn} G_j)^2 \quad (13) \]

### 2.5 Regression through the origin or with intercept?

It should be noted that there is no constant or “intercept” term included in the regression model of equation (1).

The regression goes through the origin because the geotype proportions in each time period are collectively exhaustive and mutually exclusive, so \( \sum p_{jn} = 1 \). The standard error for each coefficient is the uncertainty in the estimate for that geotype.

Alternatively, it is possible to drop one geotype from the set \( (j=1) \), and include a constant, as in equation (14).

\[ A_n = C + \sum_{j \neq 1} p_{jn} H_j + e_n \quad (14) \]

The objective function to minimise in the WLS regression is now:

\[ \sum W_n e_n^2 = \sum n W_n (A_n - C - \sum_{j \neq 1} p_{jn} H_j)^2 \quad (15) \]

The dropped geotype \( j=1 \) (perhaps the most frequent or most typical) becomes the “reference” geotype. The regression intercept \( C \) and its standard error are now the best estimate of the lump algorithm and uncertainty for the reference geotype, while each coefficient \( H_j \) estimates the difference between the reference geotype and geotype \( j \), and its standard error estimates the uncertainty in that difference.
In comparing regression through the origin with regression including an intercept, it should be noted that (provided all the non-reference geotypes are included in the intercept model):

\[ G_1 = C, \quad \text{and} \quad G_j = C + H_j, \quad j \neq 1 \]  \hspace{1cm} (16)

The standard errors of \( G_1 \) and \( C \) are in this case equal. But the standard errors of \( G_j \) and \( H_j \) \((j \neq 1)\) are not equivalent. The standard error of \( G_j \) is the uncertainty in the algorithm for geotype \( j \), while the standard error of \( H_j \) is the uncertainty in the difference between that algorithm and the reference algorithm.

The regression through the origin required that all the geotypes be included in the model. When an intercept is used, all the geotypes (except the reference geotype) can still be included in the model, even though some of them may not differ significantly from the reference geotype. Equation (16) is only valid in this case, with all non-reference geotypes explicitly included in the model. Alternatively, we can apply stepwise regression. With stepwise regression, all the geotypes are presented as potential independent variables, and will be included in order of declining significance to the model. This way, the model will include only geotypes significantly different from the norm. The most typical geotype will automatically be omitted by the tolerance or covariance test. The stepwise regression can be stopped either when the explained variance (adjusted \( R^2 \)) reaches a maximum, or until all the remaining geotypes are not significant for inclusion at some arbitrary threshold (for example, \( p>5\% \)). The intercept can then be used as the algorithm for the non-included geotypes. The intercept must be added to the appropriate coefficients to obtain the lump algorithm for each of the geotypes explicitly included in the model.

The results reported here will consider WLS regression through the origin, and with intercept. As will be discussed in the Conclusion, each form of solution may be useful for different purposes.

### 2.6 Computer programs

Weighted-least-squares (WLS) regression analysis can be carried out using a statistical package or using a spreadsheet, such as Excel.

Unfortunately, Excel’s multiple regression function LINEST caters only for OLS regression. However, Excel comes with an add-in tool called Solver, which is a very powerful general-purpose optimiser [4]. Using Solver, it is not difficult to set up a spreadsheet to do WLS regression, although error estimates are not easily obtained.

For this study, the WLS regression was done using the statistical package SPSS [5], to obtain error estimates.

### 3 Data preparation

Data for 1729 assay periods were analysed. Geologists had classified the ore into seven geotypes. For each period of operation we collected, as input data, the
kilotonnes of ore crushed and the percentage of each of the seven geotypes. From sample assays for the corresponding time periods, we compiled the lump-fines algorithm. The algorithm is defined as the lump yield, and the difference between the lump and fines assays. The first few periods of data are shown in Table 1. The \{Fe, P, Si, Al\} figures are the difference between the lump and fines assay percentages.

Table 1: Input data.

<table>
<thead>
<tr>
<th>Crush kt</th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th>G4</th>
<th>G5</th>
<th>G6</th>
<th>G7</th>
<th>Lump-Fines Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Yield</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0%</td>
<td></td>
<td></td>
<td></td>
<td>Fe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>24%</td>
<td></td>
<td></td>
<td></td>
<td>0%</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>31%</td>
<td></td>
<td></td>
<td>0%</td>
<td>Si</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100%</td>
<td>0%</td>
<td>Al</td>
</tr>
<tr>
<td>2.16</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>55% 3.08 -.022 -1.62 -1.37</td>
</tr>
<tr>
<td>3.96</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>55% 1.16 -.007 -0.74 -0.53</td>
</tr>
<tr>
<td>1.43</td>
<td>0%</td>
<td>24%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>76%</td>
<td>0%</td>
<td>54% 1.38 -.010 -0.46 -0.86</td>
</tr>
<tr>
<td>6.04</td>
<td>0%</td>
<td>31%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>69%</td>
<td>0%</td>
<td>57% 4.25 -.094 -1.11 -0.83</td>
</tr>
<tr>
<td>4.32</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>54% 3.15 -.031 -1.86 -1.76</td>
</tr>
<tr>
<td>3.04</td>
<td>0%</td>
<td>80%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>20%</td>
<td>0%</td>
<td>55% 1.65 .009 -0.69 -0.69</td>
</tr>
<tr>
<td>4.06</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>52% 0.09 -.003 -0.07 -0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4 Results

4.1 WLS regression using solver

Figure 1 is a snapshot of an Excel spreadsheet, using Solver to find the WLS regression solution for the Lump Algorithm for iron.

Following the terminology of equation (1), row 1732 contains estimates of the algorithm \( G_i \) for each of the seven geotypes. The error terms \( e_n \) are calculated for each period \( n \), in column J. Each error term is squared, multiplied by the production kilotonnes (column A), and summed into cell J1732.

The summed weighted squared error (in cell J1732) is the loss function \( \sum W_n e_n^2 \) of equation (14).

Solver was run, to minimise the value of cell J1732 by adjusting the values of the algorithm estimates in cells A1732:H1732. Solver’s solution is shown in Figure 1. It is the WLS regression solution, constrained to go through the origin.

4.2 WLS regression using SPSS

As we have seen, the spreadsheet Excel’s Solver can be used to obtain WLS regression estimates of the lump algorithm values for each geotype. However, it is not so easy to calculate standard errors of the estimates. Accordingly, the WLS regression was run again using the statistical package SPSS. The results are reported in Table 2.

The algorithm for iron is plotted in Figure 2, together with the error bar, for each geotype. It is clear that geotype G1 has a significantly larger lump-fines...
iron split, while the split for geotype G3 is significantly smaller. The remaining geotypes do not appear to be significantly different from each other.

Table 2: WLS regression result – through the origin.

<table>
<thead>
<tr>
<th>Geoype</th>
<th>Lump-Fines Algorithm - WLS Regression Solution - Through the Origin</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yield Fe P Si Al</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G1</td>
<td>51.3% +/- 2.0% 3.59 +/- 0.68 -.025 +/- .007 -1.93 +/- 0.44 -1.75 +/- 0.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G2</td>
<td>50.3% +/- 0.5% 2.10 +/- 0.17 -.020 +/- .002 -0.24 +/- 0.11 -0.21 +/- 0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G3</td>
<td>49.9% +/- 0.4% 0.26 +/- 0.15 -.006 +/- .002 0.02 +/- 0.10 -0.26 +/- 0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G4</td>
<td>53.0% +/- 0.3% 1.94 +/- 0.10 -.013 +/- .001 -1.14 +/- 0.06 -1.02 +/- 0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G5</td>
<td>52.9% +/- 0.3% 2.12 +/- 0.11 -.020 +/- .001 1.27 +/- 0.07 -1.08 +/- 0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G6</td>
<td>53.6% +/- 0.2% 1.96 +/- 0.06 -.017 +/- .001 -1.07 +/- 0.04 -0.96 +/- 0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G7</td>
<td>53.9% +/- 0.8% 2.24 +/- 0.25 -.018 +/- .003 -1.21 +/- 0.17 -1.10 +/- 0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: Solver solution for Fe lump algorithm.

Table 2: WLS regression result – through the origin.

Figure 2: The Fe lump algorithm.
4.3 Stepwise WLS regression with an intercept

The regression was repeated for each algorithm, using an intercept and stepwise regression, with a significance cut-off of \( p<5\% \).

The stepwise regression for iron yielded a constant plus significant terms for G1 and G3 only. The other five geotypes were not significant for inclusion in the algorithm model. The results are shown in Table 3.

Table 3: Stepwise WLS regression results for iron.

<table>
<thead>
<tr>
<th>Geotype</th>
<th>Fe (Lump-Fines)</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>2.00 +/- 0.04</td>
<td></td>
</tr>
<tr>
<td>Geotype G1</td>
<td>1.58 +/- 0.68</td>
<td>( p = 2.0% )</td>
</tr>
<tr>
<td>G3</td>
<td>-1.73 +/- 0.16</td>
<td>( p &lt; 0.1% )</td>
</tr>
<tr>
<td>Excluded Geotype G2</td>
<td></td>
<td>( p = 64% )</td>
</tr>
<tr>
<td>G4</td>
<td></td>
<td>( p = 57% )</td>
</tr>
<tr>
<td>G5</td>
<td></td>
<td>( p = 23% )</td>
</tr>
<tr>
<td>G6</td>
<td></td>
<td>( p = 32% )</td>
</tr>
<tr>
<td>G7</td>
<td></td>
<td>( p = 33% )</td>
</tr>
</tbody>
</table>

Table 3 shows that, for the iron algorithm, geotypes G2, G4, G5, G6, G7 all comprise a base case, each with a mean lump-fines split for iron of 2.00 +/- 0.04. For geotype G1, 1.58 +/- 0.68 should be added to the base case. For geotype G3, -1.73 +/- 0.16 should be subtracted from the base case.

It should be noted that 2.00+1.58 = 3.58 for G1, and 2.00-1.73=0.27 for G2. Within rounding error, these values agree with the estimates for G1 and G3 obtained with the “through the origin” regression of Table 2. However, the uncertainties do not compound so easily, because they have different meanings in the two cases. The standard errors in Table 2 are the uncertainties for the individual geotype estimates. The standard errors in Table 3 are the uncertainties for the differences between the geotypes and the base case.

Stepwise WLS regressions were run for the other components of the algorithm (lump yield, plus phosphorus, silica and alumina lump-fines splits). The results are collated in Table 4.

5 Conclusion

This study has demonstrated that it is not necessary to process individual geotypes through the crusher plant to establish the lump-fines algorithm for each geotype. Weighted least squares regression analysis has successfully extracted algorithm estimates for each geotype from the historically recorded data of successive periods where varying proportions of the geotypes have been crushed in each period’s production mix.
Table 4: Stepwise WLS regression – collated results.

<table>
<thead>
<tr>
<th></th>
<th>Lump-Fines Algorithm - Stepwise WLS Regression Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lump</td>
</tr>
<tr>
<td>Const</td>
<td>53.0%</td>
</tr>
<tr>
<td>G1</td>
<td>-2.6%</td>
</tr>
<tr>
<td>G2</td>
<td>-3.1%</td>
</tr>
<tr>
<td>G3</td>
<td>0.006</td>
</tr>
<tr>
<td>G4</td>
<td>0.6%</td>
</tr>
<tr>
<td>G5</td>
<td></td>
</tr>
<tr>
<td>G6</td>
<td></td>
</tr>
<tr>
<td>G7</td>
<td></td>
</tr>
</tbody>
</table>

We have seen that the algorithms can either be extracted as estimates for each geotype (even though some may not differ significantly), or as a basic reference geotype, reporting separately only those geotypes that differ significantly from the reference geotype. The first form of model is more useful to the geologist understanding the behaviour of the ore, while the second more parsimonious model is used for decision support systems, used for the daily selection of ore.

References