Institute for Polymer Research
27th Annual Symposium

Symposium documents for

Emma Daly

Abstract

Presentation

Reactivity Ratio Estimation: Statistical Issues and Solutions

E.K. Daly, T.A. Duever, A. Penlidis
Institute for Polymer Research, Department of Chemical Engineering, University of Waterloo

Reactivity ratio estimation is a nonlinear multiresponse problemwhich has been discussed extensively in the literature, due to its application to both academia and industry (Polic, 1998). Typically, reactivity ratios are estimated using the instantaneous copolymer composition equation, based on low conversion (<5%) copolymer composition data, otherwise known as the Mayo-Lewis Model. The estimation method used to determine the reactivity ratios from the Mayo-Lewis model however varies from linear techniques to non-linear, such as the error in variables model (EVM) approach. Recently sequence length distribution information, such as triad fraction data, has become of greater interest in the parameter estimation research, due to the greater number of response variables and thus potentially better estimates of the reactivity ratios. In this research the EVM parameter estimation technique is compared to the results obtained when using the standard Box Draper Determinant Criterion approach for triad fraction data. Furthermore, the potential improvement in reactivity ratio estimation using triad fraction data in place of and in addition to composition data is considered.

Equations for either composition or triad fraction data can be derived for both the terminal and penultimate models. These equations can also be further classified into either instantaneous or cumulative models. The most widely used copolymerization model is the Mayo-Lewis model (1944), which expresses the terminal model using instantaneous composition data. The terminal model assumes that only the last monomer unit on the growing chain influences the subsequent monomer addition. The Mayo-Lewis model relates the instantaneous mole fraction of monomer 1 bound in the copolymer, F_1, with the mole fractions of free monomer 1 and 2 (f_1 and f_2 respectively) via the reactivity ratios r_1 and r_2.

Whilst copolymer composition data is easily obtained and thus has been discussed extensively in the literature, the use of triad fraction data is minimal due to the experimental complications involved in obtaining this type of data in the past. However, there has been an increase in efforts to use sequence length (triad fraction) data for estimation of the reactivity ratios, given the triad fraction equations by Koenig (1980). The triad fraction equations reported by Koenig (1980) relate the instantaneous monomer 1 centered or monomer 2 centered triads to the mole fractions of free monomer 1 and 2, $f1$ and $f2$, via the reactivity ratios r_1 and r_2.

The parameter estimation techniques used to obtain reactivity ratio values, have been studied by many authors over the years including Kelen-Tudos, Fineman-Ross and Tidwell Mortimer. These authors used general linear estimation techniques and applied them to various forms of the Mayo-Lewis model using certain assumptions and manipulation of equations. However, as Rossignoli and Duever (1995) discussed, forcing the copolymerization problem into a linear form breaks the inherent assumptions of linear estimation techniques and thus the methods are statistically invalid.

Another well known estimation technique is the Box Draper Determinant Criterion (1965), which is the most popular method for multiresponse problems. However, it does require that the independent variable has insignificant error compared to the dependent variable which poses issues in the reactivity ratio estimation problem (Rossignoli and Duever, 1995). The error in variables model (EVM) is a somewhat more recent estimation technique, which does not require as many assumptions and appears to be a more suitable method for reactivity ratio estimation.

It is reasonable then to suggest that analysis of the variables of a model and their error structure is an important stage to the parameter estimation problem. Error structure refers to: the size of the error associated with each measured variable; the errors relation to the variable; and the distribution of the error. The sizes of the measurement errors in the following case studies were assumed to be 5% for feed composition and triad fraction data and 10% for copolymer composition. The errors relation to the variable is typically defined as being either additive or multiplicative. In this research we used a multiplicative error structure for composition data, as this has been reported in the literature to be a structure most applicable to the data. However, triad fraction error structure has not been studied to any great extent and thus in this research the data set is analyzed and discussed with both error structures, to illustrate the importance of understanding the error structure of the data. Lastly, the distribution assumed for this research, where error limits are only indicated, is that of uniform.

EVM versus Box Draper Determinant Criterion using Triad Fraction Data

Burke et al (1994) provided experimental NMR triad peak data for the styrene/methyl methacrylate system in bulk at 60ºC. The reactivity ratio parameter estimation problem was completed for two cases; the multiresponse determinant criterion and the error in variables model.

In the first case using the determinant criterion, the parameter estimation is a multi-response problem, where triad fractions are responses and the monomer feed fraction, f_1, is an independent variable. As discussed by Burke (1994), if all the data is used in the parameter estimation problem then issues involving co-linearity will be present, due to the triads summing to equal one. Co-linearity problems cause the solution to be very unstable and the variance of the parameter estimates to be quite large. Consequently, in order to avoid these problems two redundant variables were eliminated arbitrarily from the system.

The parameter estimation was carried out using both the multiresponse determinant criterion and the EVM approach coded in MATLAB. Comparison of the two parameter estimation methods and the effect of the error structure can be seen when the point estimates and the exact 95% confidence region for the cases are plotted (refer to Figure 1).

As can be seen the determinant criterion estimate falls outside the 95% confidence region of both the additive error and the multiplicative error EVM cases. There are two possible explanations for this behavior, the first is that the determinant criterion does not include the feed composition to be a random variable with error, but rather assumes that the error associated with the feed is negligible compared to the error in the triad fraction measurements. Furthermore, it has been shown by Oxby et al (2003) that the Box Draper Determinant Criteria method for parameter estimation is highly dependent on the data sample size.

In order to determine if the sample size was indeed the cause of the estimates not being in agreement, simulated data was used to increase the amount of data used in the analysis. Eight simulated points were used in conjunction with the experimental data provided by Burke to create a sixteen point data set. The analysis was redone using both the EVM and Box Draper Determinant Criteria methods and the resulting joint confidence regions can be seen in Figure 1 as the grey points. It can be seen that the simulated data combined with the experimental Burke (1994) data results in the estimates of the reactivity ratios being in agreement. The determinant criteria estimate using the sixteen data points falls within the JCR of both the purely experimental Burke data EVM (additive error) contour and the Simulated Burke data EVM (additive error) contour. Therefore, this shows that the Determinant Criteria estimate is highly dependent on the size of the data set, while

the EVM method (additive error) point estimate is not. As expected, the confidence region of the EVM method is reduced as the amount of data for analysis increased. Further analysis of the diagram shows that the error structure of the data greatly influences the parameter point estimates and their respective exact shaped 95% probability confidence contours.

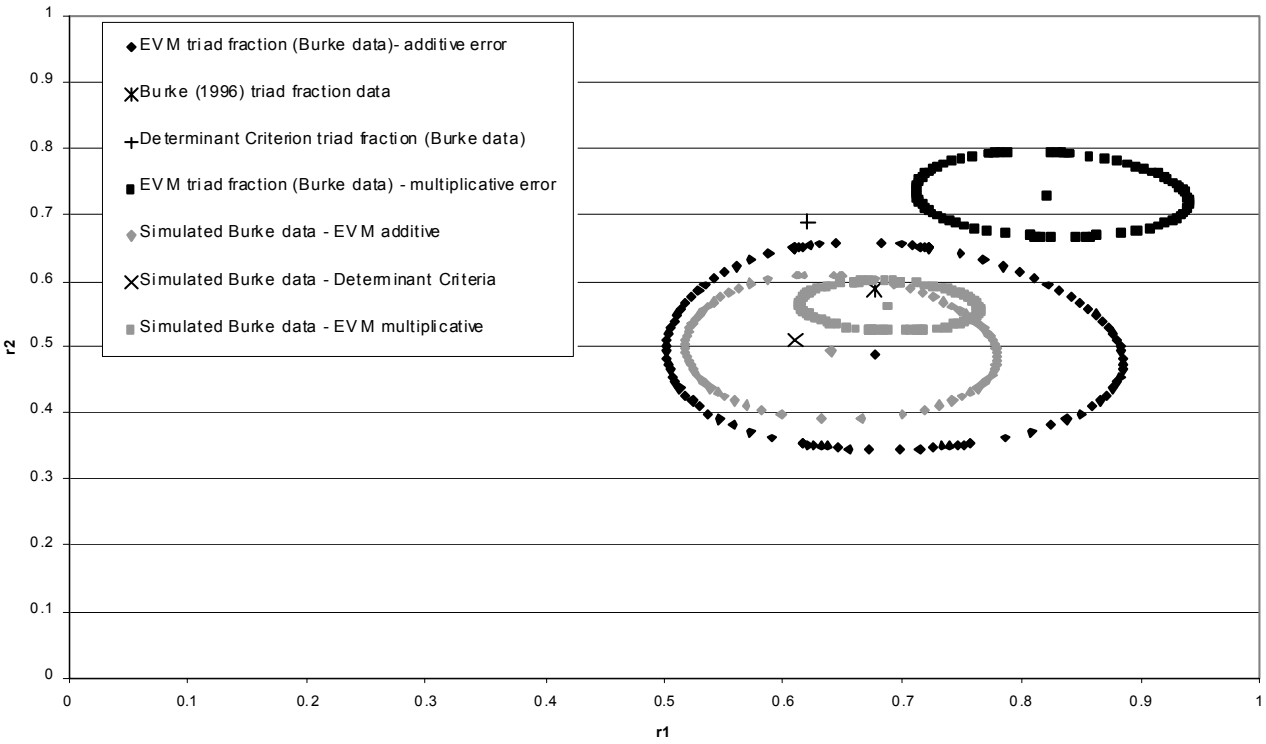


Figure 1: 95% Joint Confidence Region for EVM analysis of triad fraction data

Triad Fraction versus Composition Data using the EVM parameter estimation technique

Maxwell et al. (1993) studied the copolymerization of styrene and methyl methacrylate at 40ºC in bulk and presented extensive experimental data on both copolymer composition and triad fractions. Therefore, both data types were analyzed in this research in order to: determine whether the point estimates were influenced by the data type used; determine if one data type resulted in a smaller confidence region (less uncertainty); and evaluate the potential improvement of using both data types combined. The point estimates obtained from the different data sets along with the point estimates published by Maxwell et al (1993), can be seen in Table 1.

Table 1: Point estimates obtained using either triad fraction or composition data and the literature values published by Maxwell et al (1993).

	Data Source	Estimation Method	Error Structure	r1	r2
Maxwell (1993)	Triad fraction	NLLS		0.51	0.52
Maxwell (1993)	Composition	NLLS		0.48	0.42
Daly (2005)	Maxwell triad fraction	EVM	Additive	0.526	0.5078
Daly (2005)	Maxwell triad fraction	EVM	Multiplicative	0.6512	0.3683
Daly (2005)	Maxwell composition	EVM	Additive	0.479	0.4182
Daly (2005)	Maxwell composition	EVM	Multiplicative	0.4787	0.418
Daly (2005)	Maxwell triad and composition combined	EVM	Additive	0.5427	0.4846
Daly (2005)	Maxwell triad and composition combined	EVM	Multiplicative	0.6143	0.3683

The exact shaped 95% probability joint confidence regions of the triad fraction and the copolymer composition (refer to Figure 2) demonstrate the following:

i) The use of triad fraction data results in less uncertainty in the parameter estimates (reactivity ratios) than using the conventional copolymer composition data.

ii) The use of either multiplicative or additive error structure to copolymer composition data does not significantly affect the point estimate obtained; however the uncertainty in the parameter estimates is greatly reduced when using a multiplicative error structure.

iii) The use of either multiplicative or additive error structure to the triad fraction/NMR peak data does significantly affect the point estimates and the confidence region (uncertainty) of the parameters.

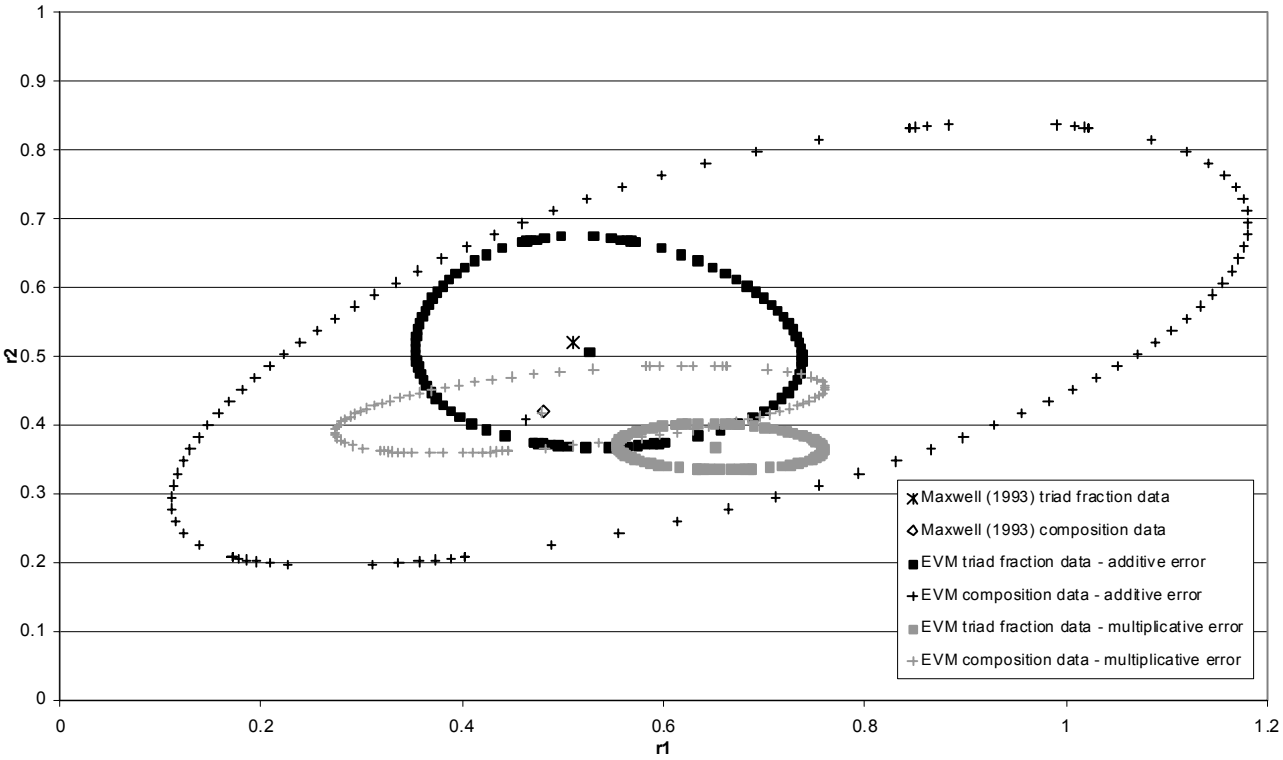


Figure 2: 95% Joint Confidence Region for EVM analysis of Maxwell et al (1993) triad data and composition data cases (additive and multiplicative error)

The question then becomes, is multiplicative error structure the incorrect structure for triad fraction/NMR normalized peak area data, or is the point estimates published in the literature incorrect due to statistically invalid parameter estimation techniques (i.e. non linear least square - NLLS)?

The next stage of analysis was to determine whether using all the available composition and triad fraction data results in a significant improvement in the degree of uncertainty of the reactivity ratio estimates compared to the confidence when using only triad fraction data (refer to Figure 3). The additional use of the copolymer composition data with the triad fraction data does not appear to have a significant increase in confidence, compared to that obtained when using only triad fraction data. That is, the slight decrease in area of the joint confidence region when both data sets are used does not seem sufficiently significant to warrant the need for conducting the measurement of the copolymer composition, regardless of which error structure is implemented.

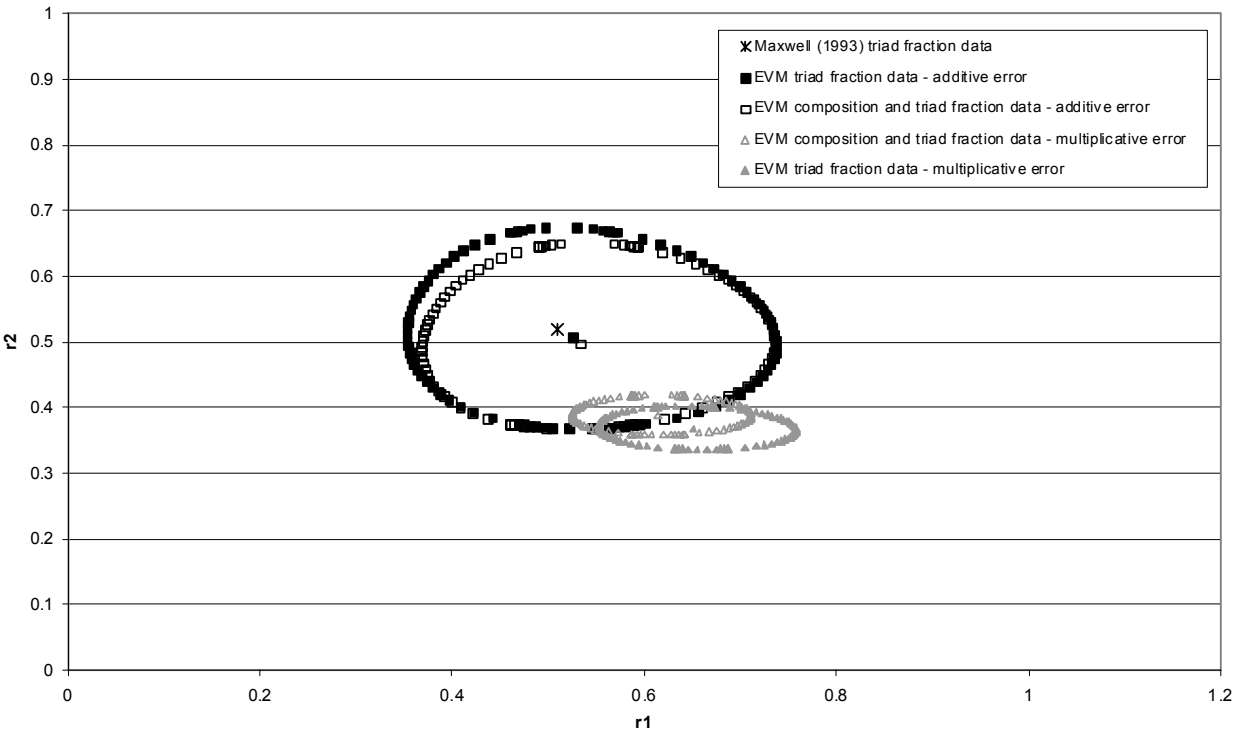


Figure 3: 95% Joint Confidence Region for EVM analysis of Maxwell et al (1994) triad and composition data combined vs. only triad fraction data

Reactivity Ratio Estimation: Statistical Issues and Solutions

Research Seminar

Emma Daly
May 18, 2005

Department of Chemical Engineering
University of Waterloo

Outline

- Introduction and objectives
- Background
  - Basic Models
  - Parameter Estimation Methods
- Results and Discussion
  - Point estimates and 95% Probability exact shaped joint confidence regions (JCR)
- Conclusions
- Future work

Introduction and Objectives

- Reactivity ratio estimation – a nonlinear parameter estimation problem.
- MATLAB program based on the EVM parameter estimation algorithm (Reilly, 1993).
- Triad fraction or composition data - which gives better estimates?
- Combining the data sets – any improvement?

Basic Models

- Terminal or Penultimate?
- Research based on Terminal Model

	Composition data		Triad Fraction data
Instantaneous Model	Mayo – Lewis Equation		Koenig Equations
Cumulative Integral Model	Analytical Solution (Meyer Lowry Model)	Numerical Solution	Numerical solution

Basic Models

	Composition data		Triad Fraction data
Instantaneous Model	**Mayo – Lewis Equation**		Koenig Equations
Cumulative Integral Model	Meyer Lowry Model	Numerical Solution	Numerical solution

$$F_1 = \frac{r_1 f_1^{\,2} + f_1 f_2}{r_1 f_1^{\,2} + 2 f_1 f_2 + r_2 f_2^{\,2}}$$

Basic Models

	Composition data		Triad Fraction data
Instantaneous Model	Mayo – Lewis Equation		**Koenig Equations**
Cumulative Integral Model	Meyer Lowry Model	Numerical Solution	Numerical solution

$$A_{111} = \frac{r_1^{\,2} f_1^{\,2}}{r_1^{\,2} f_1^{\,2} + 2 r_1 f_1 f_2 + f_2^{\,2}}, \quad A_{212} = \frac{f_2^{\,2}}{r_1^{\,2} f_1^{\,2} + 2 r_1 f_1 f_2 + f_2^{\,2}}$$

$$A_{211} = A_{112} = \frac{r_1 f_1 f_2}{r_1^{\,2} f_1^{\,2} + 2 r_1 f_1 f_2 + f_2^{\,2}} \text{ thus, } \quad A_{211+112} = \frac{2 r_1 f_1 f_2}{r_1^{\,2} f_1^{\,2} + 2 r_1 f_1 f_2 + f_2^{\,2}}$$

Basic Models

- Triad fraction data obtained from NMR peak data (Aerdts, 1993).

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0 & 0 & (1-\sigma_{12})^2 \\ 1 & (1-\sigma_{12}) & 2\sigma_{12}(1-\sigma_{12}) \\ 0 & \sigma_{12} & \sigma_{12}^{\,2} \end{bmatrix} \begin{bmatrix} A_{111} \\ A_{112+211} \\ A_{212} \end{bmatrix}$$

$$\begin{bmatrix} A \\ BC \\ D \end{bmatrix} = \begin{bmatrix} \sigma_{22} & \sigma_{22}\sigma_{12} & \sigma_{12}^{\,2} \\ 2\sigma_{22}(1-\sigma_{22}) & (1-\sigma_{12}\sigma_{22}) & (1-\sigma_{12}^{\,2}) \\ (1-\sigma_{22})^2 & 0 & 0 \end{bmatrix} \begin{bmatrix} A_{222} \\ A_{122+221} \\ A_{121} \end{bmatrix}$$

- Where σ_{22} is tacticity parameter, σ_{12} is the coisotacticity parameter (Aerdts, 1993).

Review of Estimation Methods

- Linear(ized) methods
  - Fineman-Ross (FR)
  - Kelen-Tudos (KT)
  - Extended KT
  - Inverted FR
- Nonlinear Methods
  - Nonlinear Least Squares (NLLS): Copolymer Comp. data
  - Box Draper Determinant Criterion: Triad fraction data
  - Error in Variables Model (EVM): BOTH

Review of Estimation Methods: Linear Models

- Estimates of reactivity ratios (rr) using linear models are well known to be statistically incorrect due to violation of linear regression assumptions, namely;
  - The error in the independent variable is negligible.
  - The error associated with the dependent variable is assumed independent and identically Normally distributed for the purpose of making statistical inferences.
- Linear estimates are however good initial values for NL parameter estimation problems.

Review of Estimation Methods: Nonlinear Models

- NLLS minimizes sum of squared differences between observed and fitted values of the dependent variable.
  - E.g. Mayo Lewis model: dependent variable = copolymer comp.
- Determinant Criterion minimizes determinant of the estimate of the covariance matrix.
  - Multi response problems: Multi dependent variables = triads.
- Determinant + NLLS: Negligible error in the independent variables.
  - I.e. Feed composition has insignificant error.
- EVM: Compares measured and fitted values, but it does so for all measured variables.
  - Assigns relative weights to measured quantities according to their precision.

Error Structure

- Error structure refers to:
  - size of the error associated with each measured variable.
  - errors relation to the variable (additive or multiplicative).
  - distribution of the error.
- In this work:
  - Size = 5% for feed composition & triad fraction data, 10% for copolymer composition.
  - Distribution = uniform.
  - Relation to variable = multiplicative for copolymer composition; analyzed both for triad fraction data.

Review of Estimation Methods: EVM

- EVM consists of two statements:

  1) Equating the vector of measurements X to the vector of true values ξ.

  For an additive error vector ε,

  $$X = \xi + \varepsilon$$

  while for a multiplicative error vector ε,

  $$X = \xi\,(1 + \varepsilon)$$

  2) Relates the true values of the parameters (θ^*) and variables with a model represented by;

  $$f(\xi, \theta^*) = 0$$

$$f(\xi,\theta^*) = F_1^* - \frac{r_1^* f_1^{*2} + f_1^*(1-f_1^*)}{r_1^* f_1^{*2} + 2 f_1^*(1-f_1^*) + r_2^*(1-f_1)^{*2}}$$

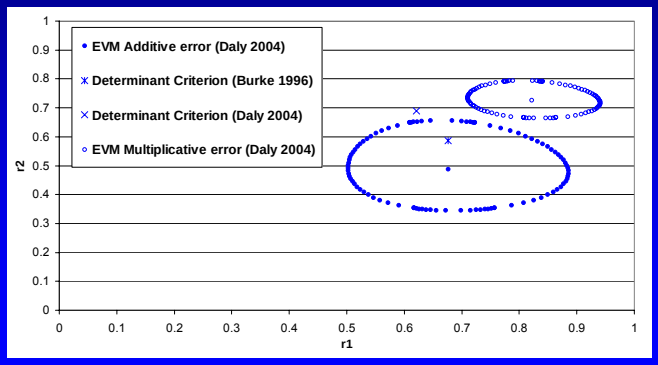
- EVM is a nested-iterative scheme (Rossignoli, 1995 or Reilly, 1993).

Triad Fraction Data: EVM vs. Determinant Criterion

- To ensure EVM program worked with triad fraction data, performed analysis on same data set using Box Draper Determinant Criterion.
- Compared the two methods performance at reactivity ratio estimation.

Results: Triad fraction data

STY-MMA: 95% Probability exact shaped Joint Confidence Regions (JCR);



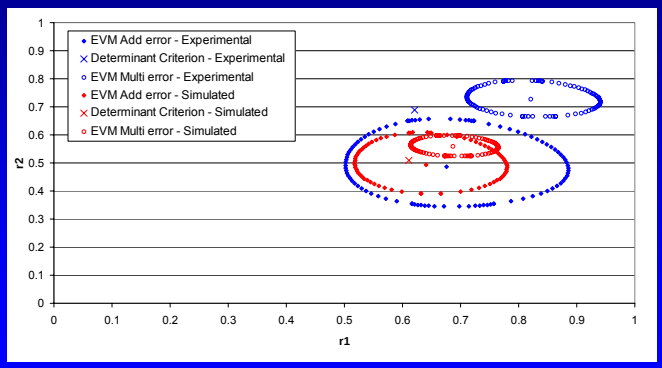
Data set provided by Burke (1997)

Discussion: EVM vs. Box Draper

- Determinant Criterion ≠ EVM
- Explanation??
  - Determinant Criterion assumes error in feed composition to be insignificant.
  - Determinant Criterion method is highly dependent on the data sample size (Oxby et al., 2003).

Results: Simulated data set

- STY-MMA: 95% Exact shaped Joint Confidence Regions (JCR)



Data set provided by Burke (1997)

Discussion

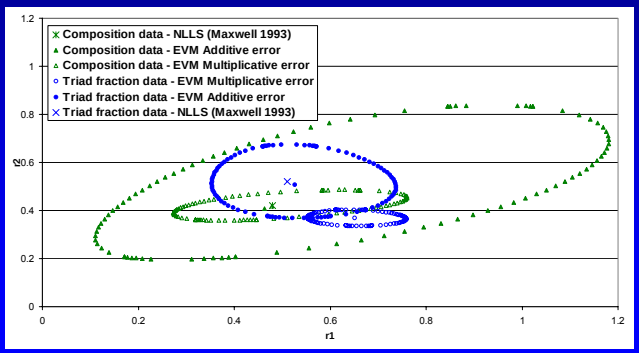
- Determinant Criterion point estimate shifts significantly when data set increased. Thus, the estimate is highly dependent on the size of the data set.
- EVM JCR area decreased as data set increased.
- Error structure influences location of the parameter point estimates.

EVM: Triad Fraction vs. Composition data

- Are triad fractions a better statistical data set for reactivity ratio parameter estimation?

EVM:Triad Fraction vs. Composition data set

- STY-MMA: 95% Exact shaped Joint Confidence Regions (JCR)



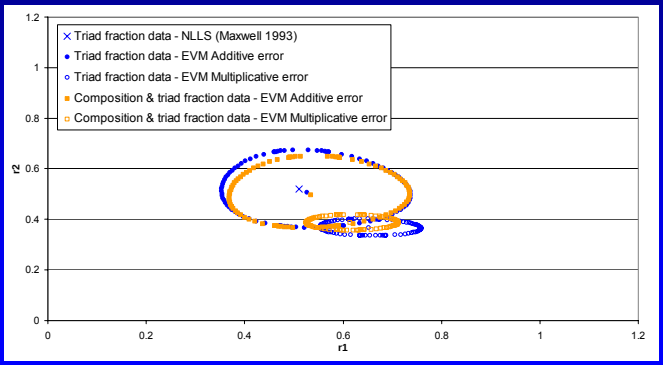
Data set provided by Maxwell et al. (1993)

Discussion

- The 95% exact shaped JCR's demonstrate the following:
  - Use of triad fraction data results in less uncertainty than using the conventional copolymer composition data.
  - Use of either error structures to copolymer composition data does not significantly affect the location of point estimates; however multiplicative structure greatly reduces uncertainty.
  - Use of either error structures to triad fraction data does significantly affect the location and confidence of the point estimates.
  - Literature value not contained in JCR of triad fraction data with a multiplicative error structure.
- Thus the question becomes: Is multiplicative error the incorrect structure for triad fraction data, or are literature point estimates incorrect?

EVM:Triad Fraction vs. Combined data set

- STY-MMA: 95% Exact shaped Joint Confidence Regions (JCR)



Data set provided by Maxwell et al. (1993)

Discussion

- Slight decrease in JCR area when both data sets are used = slight increase in confidence in the point estimates.
- Change in JCR area not sufficiently significant to warrant need for conducting the extra measurement (regardless of error structure implemented).

Conclusions

- EVM is a better statistical parameter estimation method than other conventional NL methods.
- Using triad fraction data results in less uncertainty in parameter estimates. However, NMR peak assignment and thus triad fraction data may be difficult to obtain.
- If obtaining triads are possible then no need to obtain copolymer composition data.
- Multiplicative error for triad fraction data is either incorrect structure or the literature point estimates are incorrect.

Future work

- Cumulative Composition Integral Model in EVM MATLAB program, both the analytical and numerical solution.
  - To account for feed composition drift in high conversion copolymerization.
- Terpolymer composition model in EVM MATLAB program.

Acknowledgements

- Prof. Tom Duever and Prof. Alex Penlidis
- Ontario Graduate Scholarship
- Anil Dalvi, 4th yr student who begun transfer of Fortran EVM coding to MATLAB.

Questions?