A Constrained Joint-Equation Estimation of at-a-station Hydraulic Geometry

by

Ran Shen

M.A., Simon Fraser University, 2002.

B.A., Suzhou University, 1998.

A PROJECT SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in the Department

of

Statistics and Actuarial Science

SIMON FRASER UNIVERSITY

Spring 2005

APPROVAL

Name: Ran Shen

Degree: Master of Science

Title of project: A Constrained Joint-Equation Estimation of at-a-station Hydraulic Geometry

Examining Committee: Dr. Gary Parker
Chair

____

Dr. Carl Schwarz
Senior Supervisor
Simon Fraser University

____

Dr. Richard Lockhart
Simon Fraser University

____

Dr. Joan Hu
External Examiner
Simon Fraser University

Date Approved: ______________________________

Abstract

Hydraulic geometry describes the relations between a stream's discharge and its width, depth and velocity as a system. The continuity equation implies two cross-equation constraints, one of which is that the hydraulic exponents should sum to one. Traditional literature, using one-at-a-time estimation methods, have either ignored the constraints, or force the exponents to sum to unity by arbitrarily manipulating the estimates. By using a systematic approach called Seemingly Unrelated Regression(SUR), we are able to jointly estimate the relations and impose the constraints. The unrestricted and restricted estimates are computed from SUR method and Ordinary Least Squares(OLS). It is found that SUR and OLS yield identical unrestricted results when the sets of regressors are identical. Although SUR estimates are asymptotically at least as efficient as OLS, due to the finite sample sizes of our data, the restricted estimates from SUR generally have larger standard errors than the restricted OLS.

Dedication

To my parents.

Acknowledgements

This work would not have been possible without my supervisor Dr. Carl Schwarz. Although he is an excellent professor and supervisor, I am not necessarily a good student. I cannot thank him enough for his patience, encouragement, guidance and support.

I also wish to acknowledge Scott Babakaiff at the Ministry of Water, Land and Air Protection, for bringing up the issue of placing joint constraints on hydraulic geometry estimates, and for collecting and providing all the data.

I am especially indebted to Dr. Richard Lockhart, who was the then-Graduate Chair when I was admitted, for giving me a chance to pursue a career in statistics. (Sorry Richard that I only looked good on paper.) Richard also taught my very first two graduate courses in statistics. His help was essential to my survival of the graduate school.

Special thanks to Steve Overduin, who has offered me numerous (and free) ethic counseling and spiritual support. To Matthew Pratola and Pritam Ranjan, who always get on my nerves by poking fun at me, for giving me extra motivation to get out of the program. To Wilson Lu, who was a tremendous help during my first year. To my dear friends Frances, Alice and Grace, for being there to share my laughs, and occasionally my tears. To Chunfang Lin, Darcy Pickard, Suman Jiwani, Jean Shin, Maria Lorenzi, Amy Summers, Jason Nielson, Simon Bonner, John Bentley, and Eric Sayre, for helping me out with courses, programming and English, for the many fond memories we share (how can I ever forget the Grouse Grind and our annual retreat), for the dynamite conversation that always goes on in K9501, and for so much more.

Finally, to the faculty, staff and graduate students in the Department of Statistics and Actuarial Science, if life is a journey, I'd like to thank you all for giving me such a wonderful ride.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Leopold and Maddock (1953) describe the relationships between a stream channel's width, depth, velocity and its discharge by three power functions:

$$
\begin{aligned}
w &= aQ^b \\
d &= cQ^f \qquad (1.1) \\
v &= kQ^m
\end{aligned}
$$

where w is the wetted width; d is the mean depth, defined as the depth of a rectangular channel having the same surface width and cross-sectional area as the actual stream; v is the mean velocity, defined as the discharge per cross-sectional area of the stream; Q is the discharge or flow; and a, c, k, b, f, m are the parameters of the relationship.

These relationships are called hydraulic geometry, which has been used extensively to describe the flow behavior of natural rivers in a variety of regions. Regional hydraulic geometry curves are developed to help regulatory agencies with river morphology assessment and instream ecology and river management. It "provides a promising method for making an initial assessment of environmental impacts of proposed flow changes" (Babakaiff, 2004).

Hydraulic geometry is measured in two ways: 'at-a-station' where discharge is measured over time at a given site, or 'downstream' where discharge is measured at various sites along a stream at a given frequency. The parameters b, f, m are called hydraulic exponents. The continuity equation

$$Q = wdv \tag{1.2}$$

implies two constraints on the parameters: $b + f + m = 1$, which means hydraulic exponents are unit-sum constrained, and $ack = 1$.

One way to estimate the parameters a, c, k, b, f, m is to use a log-linear model (LLM). If we take a log transformation of the variables in system (1.1) we obtain three linear functions:

$$\begin{aligned} \log w &= \alpha_w + \beta_w \log Q \\ \log d &= \alpha_d + \beta_d \log Q \\ \log v &= \alpha_v + \beta_v \log Q \end{aligned} \tag{1.3}$$

In terms of the LLM, the two constraints become

$$\beta_w + \beta_d + \beta_v = 1 \tag{1.4}$$

$$\alpha_w + \alpha_d + \alpha_v = 0 \tag{1.5}$$

However, in the presence of a nonlinear relationship between channel roughness and discharge, the LLM can no longer describe hydraulic geometry sufficiently. Richards (1973) proposed a log quadratic model (LQM) defined by

$$\begin{aligned} \log w &= \beta_1 + \beta_2 \log Q + \beta_3 (\log Q)^2 \\ \log d &= f_1 + f_2 \log Q + f_3 (\log Q)^2 \\ \log v &= m_1 + m_2 \log Q + m_3 (\log Q)^2 \end{aligned} \tag{1.6}$$

By taking the first-order derivative of system (1.6) with respect to log Q, he incorrectly concluded that the continuity equation (1.2) implies

$$\beta_1 + f_1 + m_1 = 0 \tag{1.7}$$

$$\beta_2 + f_2 + m_2 + 2(\beta_3 + f_3 + m_3)(\log Q) = 1 \tag{1.8}$$

so that hydraulic exponents are allowed to be negative. The correct set of constraints are

$$\beta_1 + f_1 + m_1 = 0$$

$$\beta_2 + f_2 + m_2 = 1$$

$$\beta_3 + f_3 + m_3 = 0$$

Brush (1961) found that four streams in the Appalachians have negative velocity exponents, which implies a downstream decrease in velocity. Fortunately, Richards (1973) pointed out that nonlinear systems occur only occasionally. Rhodes (1977) recommended a screening procedure which rejects hydraulic exponents whose sum is outside the range 0.95 - 1.05, and Ridenour and Giardino (1991) suggested using this procedure to eliminate highly nonlinear systems.

The current method of estimating the LLM model is to run three separate least-squares regression after the logarithmic transformation. More often than not, the unit-sum constraints specified in equation (1.4) and (1.5) are not satisfied. For example, Rhodes (1977) reported that out of 332 sets of equations inspected, only 125 sets of exponents summed to unity, and 17 sets had reported exponents whose sum was outside the range 0.95 - 1.05. In addition, no consideration seems to be given to the constraint specified in equation (1.5).

Departure from the unit-sum constraint has been handled in two ways. Park (1977) apportioned the deviation equally among the three exponents, i.e., if the hydraulic exponents sum to 0.9, then each value is divided by 0.9 such that their sum would be forced to 1. Rhodes (1977) excluded deviations greater than 0.05, and adjusted the values of β_2 and β_3 since width is subject to the least measurement error. The adjustments were made by adding or subtracting values from β_2 and β_3 according to a table he presented. For example, if the exponents sum to 1.05, then β_2 will drop by 0.02 and β_3 will drop by 0.03. If the sum is 1.04, then β_2 and β_3 will both go down by 0.02.

These two *ad hoc* approaches to treat departure from unity are problematic. They are incapable of estimating the standard errors of the estimates, nor can they test for the unit-sum constraint. Furthermore, they fail to account for the different precisions of the individual estimates. Rhodes (1977) tried to solve this problem by adjusting only the exponents for depth and velocity, but he also conceded that the adjusted values were only close estimates of the true parameters, and it is difficult to tell whether the adjustment were made in the right direction. Finally, both methods cannot be used to make inverse predictions to forecast discharge from given values of width, depth or velocity.

This paper is concerned with only at-a-station hydraulic geometry. Chapter 2 discusses a joint-equation estimation method, which allows for estimating a system of three equations with two restrictions. The method is compared to OLS and illustrated with numerical examples in Chapter 3. Finally, a brief summary is provided in Chapter 4.

Chapter 2

A Systematic Approach

This chapter covers the following ground:

1. Review of restricted and un-restricted single-equation estimation. Section(2.1) discusses the estimate under the assumption of spherical disturbances. The assumption will be relaxed in Section(2.2).

2. Review of constraint testing in Section(2.3).

3. Restricted and un-restricted joint-equation estimation. Section(2.4) shows the SUR method to simultaneously estimate equations and to impose restrictions across equations.

4. Inverse prediction to find inverse confidence intervals in Section(2.5).

2.1 Single-Equation Constrained OLS Estimate

Consider first the usual unconstrained OLS estimation. For a classical linear regression model $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$ where $V(\boldsymbol{y}|\boldsymbol{X}) = \sigma^2\boldsymbol{I}$ ($\sigma^2 > 0$), OLS estimation is equivalent to minimizing the following function of $\boldsymbol{\beta}$

$$f(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta}) = \boldsymbol{y'y} - 2\boldsymbol{y'X\beta} + \boldsymbol{\beta'X'X\beta}$$

The derivative of $f(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ is

$$f'(\boldsymbol{\beta}) = -2\boldsymbol{X'y} + 2\boldsymbol{X'X\beta}$$

This leads to the familiar OLS solution that $\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'y}$ with $V(\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}}) = (\boldsymbol{X'X})^{-1}\sigma^2$. It is the best unbiased estimator when the errors are normally distributed.

Since $\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}}$ minimizes the residual sum of squares, any other estimator, say $\hat{\boldsymbol{\eta}}$, must yield a larger sum of squares. The disturbance vector associated with $\hat{\boldsymbol{\eta}}$ is

$$\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\eta}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}} - \boldsymbol{X}(\hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})$$

and its sum of squares

$$\begin{aligned}
&(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\eta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\eta}}) \\
&= (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}}) + (\hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})'\boldsymbol{X'X}(\hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\beta}}_{\boldsymbol{ols}}) \\
&-2(\hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})'\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}}) \\
&= (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}}) + (\hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})'\boldsymbol{X'X}(\hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})
\end{aligned} \tag{2.1}$$

as $\boldsymbol{X}$ is orthogonal to $(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})$. Thus the excess sum of squares associated with $\hat{\boldsymbol{\eta}}$ is a positive definite quadratic form with $\boldsymbol{X'X}$ as its matrix and $\hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\beta}}_{\boldsymbol{ols}}$ as its vector.

It will be assumed that $\boldsymbol{y}$ is of order $n \times 1$ and $\boldsymbol{X}$ is a full-rank $n \times K$ matrix. Now suppose the arguments of $f(\boldsymbol{\beta})$ are subject to certain linear constraints

$$\gamma = \boldsymbol{R\beta}$$

where $\boldsymbol{\gamma}$ and $\boldsymbol{R}$ are given matrices of order $q \times 1$ and $q \times K$, respectively. The constrained estimate can be obtained by constructing a function $F(\boldsymbol{\beta}, \boldsymbol{\phi})$:

$$F(\boldsymbol{\beta}, \boldsymbol{\phi}) = (\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta}) - \boldsymbol{\phi}'(\gamma - \boldsymbol{R\beta})$$

By differentiating $F(\boldsymbol{\beta}, \boldsymbol{\phi})$ with $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$ and setting the derivatives equal to zero, we obtain the following results:

$$\frac{\partial F}{\partial \boldsymbol{\beta}} = -2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X\beta} + \boldsymbol{R}'\boldsymbol{\phi} = 0 \tag{2.2}$$

$$\frac{\partial F}{\partial \boldsymbol{\phi}} = -(\gamma - \boldsymbol{R\beta}) = 0 \tag{2.3}$$

To get an estimate of $\boldsymbol{\phi}$, we pre-multiply both sides of equation (2.2) by $\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}$:

$$-2\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X\beta} + \boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\boldsymbol{\phi} = 0$$

Since $\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X\beta} = \boldsymbol{R\beta} = \gamma$ is the constraint, and $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ is the unconstrained estimate $\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}}$,

$$\hat{\boldsymbol{\phi}} = -2[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\gamma - \boldsymbol{R}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}}) \tag{2.4}$$

The constrained OLS estimate $\hat{\boldsymbol{\beta}}^*_{\boldsymbol{ols}}$ is then found by pre-multiplying both sides of equation(2.2) by $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ and replacing $\boldsymbol{\phi}$ with the right hand side of equation(2.4):

$$\hat{\boldsymbol{\beta}}^*_{\boldsymbol{ols}} = \hat{\boldsymbol{\beta}}_{\boldsymbol{ols}} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\gamma - \boldsymbol{R}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}}) \tag{2.5}$$

The difference between $\hat{\boldsymbol{\beta}}^*_{\boldsymbol{ols}}$ and $\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}}$ is linear in the vector $(\boldsymbol{\gamma} - \boldsymbol{R}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})$, and it measures the deviance of $\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}}$ from the constraint. Hence, if $\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}}$ happens to satisfy the constraints $\boldsymbol{\gamma} = \boldsymbol{R\beta}$, then $\hat{\boldsymbol{\beta}}^*_{\boldsymbol{ols}}$ and $\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}}$ are identical.

$\hat{\boldsymbol{\beta}}^*_{\boldsymbol{ols}}$ is unbiased, for it can be written as follows:

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^*_{\boldsymbol{ols}} &= \boldsymbol{\beta} + (\boldsymbol{X'X})^{-1}\boldsymbol{X'\epsilon} + (\boldsymbol{X'X})^{-1}\boldsymbol{R'}[\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{R'}]^{-1} \\
&(\boldsymbol{R\beta} - \boldsymbol{R\beta} - \boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{X'\epsilon}) \\
&= \boldsymbol{\beta} + [(\boldsymbol{X'X})^{-1} - (\boldsymbol{X'X})^{-1}\boldsymbol{R'}[\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{R'}]^{-1}\boldsymbol{R}(\boldsymbol{X'X})^{-1}]\boldsymbol{X'\epsilon}
\end{aligned}$$

Based on equations (2.1) and (2.5), the excess sum of squares associated with $\hat{\boldsymbol{\beta}}^*_{\boldsymbol{ols}}$

$$\begin{aligned}
&(\hat{\boldsymbol{\beta}}^*_{\boldsymbol{ols}} - \hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})'\boldsymbol{X'X}(\hat{\boldsymbol{\beta}}^*_{\boldsymbol{ols}} - \hat{\boldsymbol{\beta}}^*_{\boldsymbol{ols}}) \\
&= (\gamma - \boldsymbol{R}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})'[\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{R'}]^{-1}\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{X'X}(\boldsymbol{X'X})^{-1}\boldsymbol{R'} \\
&[\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{R'}]^{-1}(\gamma - \boldsymbol{R}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}}) \\
&= (\gamma - \boldsymbol{R}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})'[\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{R'}]^{-1}(\gamma - \boldsymbol{R}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})
\end{aligned} \tag{2.6}$$

is a positive definite quadratic form with $[\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{R'}]^{-1}$ as its matrix and $(\gamma - \boldsymbol{R}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})$ as its vector. This excess is caused by the linear constraints.

The variance of $\hat{\boldsymbol{\beta}}^*_{\boldsymbol{ols}}$ is

$$V(\hat{\boldsymbol{\beta}}^*_{\boldsymbol{ols}}) = (\boldsymbol{X'X})^{-1}\sigma^2 - (\boldsymbol{X'X})^{-1}\boldsymbol{R'}[\boldsymbol{R}(\boldsymbol{X'X})^{-1}\boldsymbol{R'}]^{-1}\boldsymbol{R}(\boldsymbol{X'X})^{-1}\sigma^2 \tag{2.7}$$

Clearly, $V(\hat{\boldsymbol{\beta}}^*_{\boldsymbol{ols}})$ is smaller than $V(\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})$ by a positive semidefinite matrix. The variance reduction is the precision gained from imposing the constraint. Proof of equation(2.7) is left to the next section.

2.2 Single-Equation Constrained GLS Estimate

When the conditional variance matrix of $\boldsymbol{y}$ given $\boldsymbol{X}$ is not scalar, i.e., $V(\boldsymbol{y}|\boldsymbol{X}) = \sigma^2\boldsymbol{V}$, where $\boldsymbol{V}$ is a symmetric positive definite $n \times n$ matrix, $\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}}$ is still unbiased, but it no longer has minimum variance among all linear unbiased estimators.

To find an estimator that is best in this situation, consider transforming the disturbance vector ϵ such that its covariance matrix after transformation is scalar again. As $\boldsymbol{V}$ is symmetric and positive definite, there exists an $n \times n$ nonsingular matrix $\boldsymbol{P}$ such that $\boldsymbol{P'P = V^{-1}}$. It follows that $E(\boldsymbol{P\epsilon}) = 0$ and $V(\boldsymbol{P\epsilon}) = \boldsymbol{P}V(\boldsymbol{\epsilon})\boldsymbol{P'} = \sigma^2\boldsymbol{PVP'} = \sigma^2\boldsymbol{I}$.

The best linear unbiased estimator is the solution that minimizes

$$
\begin{aligned}
f(\boldsymbol{\beta}) &= \boldsymbol{(P\epsilon)'(P\epsilon)} = \boldsymbol{(Py - PX\beta)'(Py - PX\beta)} = \boldsymbol{(y - X\beta)'P'P(y - X\beta)} \\
&= \boldsymbol{(y - X\beta)'V^{-1}(y - X\beta) = y'V^{-1}y - 2y'V^{-1}X\beta + \beta'X'V^{-1}X\beta}
\end{aligned}
$$

The derivative of $f(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ is

$$
f'(\boldsymbol{\beta}) = \boldsymbol{-2X'V^{-1}y + 2X'V^{-1}X\beta}
$$

Hence we obtain the GLS estimate

$$
\hat{\boldsymbol{\beta}}_{\boldsymbol{gls}} = \boldsymbol{(X'V^{-1}X)^{-1}X'V^{-1}y} \tag{2.8}
$$

and its variance

$$
V(\hat{\boldsymbol{\beta}}_{\boldsymbol{gls}}) = \sigma^2\boldsymbol{(X'P'PX)}^{-1} = \sigma^2(\boldsymbol{X'V^{-1}X})^{-1} \tag{2.9}
$$

If we incorrectly assume the errors are i.i.d when they are not, applying OLS has the following consequences:

1. $\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}}$ is unbiased, but inefficient relative to $\hat{\boldsymbol{\beta}}_{\boldsymbol{gls}}$.

2. The variance of $\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}}$ is $\sigma^2(\boldsymbol{X'X})^{-1}\boldsymbol{X'VX}(\boldsymbol{X'X})^{-1}$ instead of $\sigma^2(\boldsymbol{X'X})^{-1}$.

3. The OLS estimator of σ^2, $(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})'(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})/(n-K)$, is biased.

If the same linear restrictions $\boldsymbol{\gamma}=\boldsymbol{R\beta}$ is imposed, the constrained GLS estimate is the solution that minimizes

$$F(\boldsymbol{\beta},\boldsymbol{\phi})=(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta})-\boldsymbol{\phi}'(\gamma-\boldsymbol{R\beta})$$

over β. Again we differentiate $F(\boldsymbol{\beta},\boldsymbol{\phi})$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$:

$$\frac{\partial F}{\partial \boldsymbol{\beta}}=-2\boldsymbol{X'V}^{-1}\boldsymbol{y}+2\boldsymbol{X'V}^{-1}\boldsymbol{X\beta}+\boldsymbol{R'\phi}=0 \tag{2.10}$$

$$\frac{\partial F}{\partial \boldsymbol{\phi}}=-(\gamma-\boldsymbol{R\beta})=0 \tag{2.11}$$

To obtain $\hat{\boldsymbol{\phi}}$, we pre-multiply both sides of equation (2.10) by $\boldsymbol{R}(\boldsymbol{X'V}^{-1}\boldsymbol{X})^{-1}$, and replace $(\boldsymbol{X'V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X'V}^{-1}\boldsymbol{y}$ with $\hat{\boldsymbol{\beta}}_{\boldsymbol{gls}}$, $\boldsymbol{R\beta}$ with γ:

$$\hat{\boldsymbol{\phi}}=-2[\boldsymbol{R}(\boldsymbol{X'V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\gamma-\boldsymbol{R}\hat{\boldsymbol{\beta}}_{\boldsymbol{gls}}) \tag{2.12}$$

The constrained GLS estimate $\hat{\boldsymbol{\beta}}^*_{\boldsymbol{gls}}$ is found by pre-multiplying both sides of equation(2.10) by $(\boldsymbol{X'V}^{-1}\boldsymbol{X})^{-1}$, and substituting the right hand side of equation(2.12) for $\boldsymbol{\phi}$. Thus,

$$\hat{\boldsymbol{\beta}}^*_{\boldsymbol{gls}}=\hat{\boldsymbol{\beta}}_{\boldsymbol{gls}}+(\boldsymbol{X'V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X'V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\gamma-\boldsymbol{R}\hat{\boldsymbol{\beta}}_{\boldsymbol{gls}}) \tag{2.13}$$

Let $\boldsymbol{C}=(\boldsymbol{X'V}^{-1}\boldsymbol{X})^{-1}$,

$$\begin{aligned}\hat{\boldsymbol{\beta}}^*_{\boldsymbol{gls}}&=\hat{\boldsymbol{\beta}}_{\boldsymbol{gls}}+\boldsymbol{CR}'(\boldsymbol{RCR}')^{-1}(\gamma-\boldsymbol{R}\hat{\boldsymbol{\beta}}_{\boldsymbol{gls}})\\&=\boldsymbol{\beta}+\boldsymbol{CX'V}^{-1}\boldsymbol{\epsilon}+\boldsymbol{CR}'(\boldsymbol{RCR}')^{-1}(\boldsymbol{R\beta}-\boldsymbol{R\beta}-\boldsymbol{RCX'V}^{-1}\boldsymbol{\epsilon})\\&=\boldsymbol{\beta}+[\boldsymbol{C}-\boldsymbol{CR}'(\boldsymbol{RCR}')^{-1}\boldsymbol{RC}]\boldsymbol{X'V}^{-1}\boldsymbol{\epsilon}\end{aligned}$$

Obviously, $\hat{\boldsymbol{\beta}}^*_{\boldsymbol{gls}}$ is also unbiased. Its variance is

$$\boldsymbol{V}(\hat{\boldsymbol{\beta}}^*_{\boldsymbol{gls}})=\sigma^2[\boldsymbol{C}-\boldsymbol{CR}'(\boldsymbol{RCR}')^{-1}\boldsymbol{RC}] \tag{2.14}$$

To prove equation(2.14), let $\boldsymbol{T} = \boldsymbol{C} - \boldsymbol{C}\boldsymbol{R}'(\boldsymbol{R}\boldsymbol{C}\boldsymbol{R}')^{-1}\boldsymbol{R}\boldsymbol{C}$,

$$\begin{aligned} V(\hat{\boldsymbol{\beta}}^*_{gls}) &= E[(\hat{\boldsymbol{\beta}}^*_{gls} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}^*_{gls} - \boldsymbol{\beta})'] \\ &= E[(\boldsymbol{T}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{\epsilon})(\boldsymbol{T}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{\epsilon})'] \\ &= \boldsymbol{T}\boldsymbol{X}'\boldsymbol{V}^{-1}(\sigma^2\boldsymbol{V})\boldsymbol{V}^{-1}\boldsymbol{X}\boldsymbol{T}' = \sigma^2[\boldsymbol{C} - \boldsymbol{C}\boldsymbol{R}'(\boldsymbol{R}\boldsymbol{C}\boldsymbol{R}')^{-1}\boldsymbol{R}\boldsymbol{C}] \end{aligned}$$

after some manipulation. Equation(2.7) is a special case of equation(2.14) where $\boldsymbol{V}$ is replaced by $\boldsymbol{I}$.

The estimator $\hat{\boldsymbol{\beta}}_{gls}$ does not minimize the residual sum of squares, but it minimizes the sum of squares of the transformed disturbance vector $\boldsymbol{P}\boldsymbol{\epsilon}$. Any other estimator $\hat{\boldsymbol{\eta}}$ leads to an excess sum of squares, which is equal to $(\hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\beta}}_{gls})'\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}(\hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\beta}}_{gls})$. In a similar fashion to how we proved equation(2.6), the excess sum of squares associated with $\hat{\boldsymbol{\beta}}^*_{gls}$ is

$$(\boldsymbol{r} - \boldsymbol{R}\hat{\boldsymbol{\beta}}_{gls})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{r} - \boldsymbol{R}\hat{\boldsymbol{\beta}}_{gls}) \tag{2.15}$$

2.3 Testing the Constraint

2.3.1 Under OLS

The likelihood ratio of the unconstrained maximum of the likelihood function to the constrained maximum is

$$\begin{aligned} &[\frac{(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta_{ols}})'(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta_{ols}})}{(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta^*_{ols}})'(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta^*_{ols}})}]^{n/2} \\ =&[1+\frac{(\boldsymbol{r}-\boldsymbol{R}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{r}-\boldsymbol{R}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})}{(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})'(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})}]^{-n/2} \end{aligned} \tag{2.16}$$

where the second line makes use of equation(2.1) and equation(2.6).

If the constraints hold, $\gamma-\boldsymbol{R}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}}=-\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\epsilon}$. Thus

$$\begin{aligned} &(\gamma-\boldsymbol{R}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\gamma-\boldsymbol{R}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}}) \\ &=\boldsymbol{\epsilon}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\epsilon} \end{aligned}$$

is a quadratic form with $\boldsymbol{\epsilon}$ as the vector. Its matrix

$\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$

is idempotent with rank q.

Therefore, if the null hypothesis is true, $(\gamma-\boldsymbol{R}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\gamma-\boldsymbol{R}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})$ is distributed as $\sigma^2\chi^2_q$. The denominator $(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})'(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})$ can also be written as a quadratic form $\boldsymbol{y}'\boldsymbol{M}\boldsymbol{y}$ or $\boldsymbol{\epsilon}'\boldsymbol{M}\boldsymbol{\epsilon}$, where $\boldsymbol{M}=[\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']$ is another idempotent matrix with rank $n-K$. The denominator is distributed as $\sigma^2\chi^2_{n-K}$. In fact, it equals $(n-K)s^2$ under the assumptions of a standard linear model.

The two quadratic forms are independent, because the product of their matrices is 0. Hence the test statistic

$$\frac{(\gamma-\boldsymbol{R}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\gamma-\boldsymbol{R}\hat{\boldsymbol{\beta}}_{\boldsymbol{ols}})/q}{s^2} \tag{2.17}$$

is distributed as $F(q,n-K)$.

2.3.2 Under GLS

Under GLS, the likelihood ratio is constructed as

$$\begin{aligned}&[\frac{(\boldsymbol{y}-\boldsymbol{X\beta_{gls}})'\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta_{gls}})}{(\boldsymbol{y}-\boldsymbol{X\beta^*_{gls}})'\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta^*_{gls}})}]^{n/2}\\ =&[1+\frac{(\gamma-\boldsymbol{R\hat{\beta}_{gls}})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\gamma-\boldsymbol{R\hat{\beta}_{gls}})}{(\boldsymbol{y}-\boldsymbol{X\hat{\beta}_{gls}})'\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\hat{\beta}_{gls}})}]^{-n/2}\end{aligned} \tag{2.18}$$

The denominator is also distributed as $\sigma^2\chi^2_{n-K}$:

$$\begin{aligned}&(\boldsymbol{y}-\boldsymbol{X\hat{\beta}_{gls}})'\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\hat{\beta}_{gls}})\\ &=[\boldsymbol{X\beta}+\boldsymbol{\epsilon}-\boldsymbol{X\beta}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{\epsilon}]'\boldsymbol{V}^{-1}[\boldsymbol{X\beta}+\boldsymbol{\epsilon}-\boldsymbol{X\beta}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{\epsilon}]\\ &=\boldsymbol{\epsilon}'[\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}]'\boldsymbol{V}^{-1}[\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}]\boldsymbol{\epsilon}\\ &=\boldsymbol{\epsilon}'[\boldsymbol{V}^{-1}-\boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}]\boldsymbol{\epsilon}\\ &=(\boldsymbol{P\epsilon})'[\boldsymbol{I}-\boldsymbol{PX}(\boldsymbol{X}'\boldsymbol{P}'\boldsymbol{PX})^{-1}\boldsymbol{X}'\boldsymbol{P}'](\boldsymbol{P\epsilon})\end{aligned} \tag{2.19}$$

Although the $\boldsymbol{\epsilon}$'s are not i.i.d, after the transformation the $\boldsymbol{P\epsilon}$'s are standard normal covariates. $(\boldsymbol{y}-\boldsymbol{X\hat{\beta}_{gls}})'\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\hat{\beta}_{gls}})$ is again a quadratic form. Its matrix $\boldsymbol{I}-\boldsymbol{PX}(\boldsymbol{X}'\boldsymbol{P}'\boldsymbol{PX})^{-1}\boldsymbol{X}'\boldsymbol{P}'$ is idempotent of rank $n-K$. Its vector $\boldsymbol{P\epsilon}$ has a scalar covariance matrix $\sigma^2\boldsymbol{I}$, as shown earlier in section(2.2).

If the null hypothesis is true, $\boldsymbol{\gamma}-\boldsymbol{R\hat{\beta}_{gls}}$ can be written as $-\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{P}'\boldsymbol{P\epsilon}$, therefore

$$\begin{aligned}&(\gamma-\boldsymbol{R\hat{\beta}_{gls}})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\gamma-\boldsymbol{R\hat{\beta}_{gls}})\\ &=-\boldsymbol{\epsilon}'\boldsymbol{P}'\boldsymbol{PX}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(-\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{P}'\boldsymbol{P\epsilon})\\ &=(\boldsymbol{P\epsilon})'\boldsymbol{PX}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{P}'(\boldsymbol{P\epsilon})\end{aligned} \tag{2.20}$$

is another quadratic form distributed as $\sigma^2\chi_q^2$.

The two quadratic forms are independent, since it can be proven that the product of their matrices equals 0. Hence, the test statistic

$$\frac{n-K}{q} \times \frac{(\gamma - \boldsymbol{R}\hat{\boldsymbol{\beta}}_{\boldsymbol{gls}})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\gamma - \boldsymbol{R}\hat{\boldsymbol{\beta}}_{\boldsymbol{gls}})}{(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{gls}})'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{gls}})} \tag{2.21}$$

is distributed as $F(q, n-K)$.

2.4 Seemingly Unrelated Regression Estimate

So far, we have restricted our discussions to the single-equation case. But the hydraulic geometry problem requires estimating simultaneously the complete sets of parameters of the three equations in a system, and placing two constraints on coefficients across equations.

Zellner(1962) developed a SUR method (Seemingly Unrelated Regression) to algebraically represent a multi-equation model in a single-equation form. It is a generalization of OLS for a multivariate regression model, and produces consistent and asymptotically efficient estimates for systems of regression equations. SUR refers to the fact that although the equations appear unrelated, they are in fact connected by correlations of the disturbances. An example discussed extensively was the application of Grunfeld's investment model (Grunfeld, 1958) to two firms in the same industry. Suppose real gross investment of a firm is determined by its market value and capital stock at the beginning of a period. It is reasonable to assume that common market forces will influence both firms, and likely the errors of the two regressions are correlated. Therefore, rather than running two separate regressions, it makes sense to treat them as a system and take into account the possible correlation among the disturbances. For multivariate Gaussian response, SUR has become a well-established procedure in econometrics.

We are interested in using SUR because system(1.1) is a simplified summary of the complicated relations between discharge and channel characteristics. Many other factors, such as channel size, shape and slope, are considered implicitly. Therefore, the disturbances across equations are no longer independent.

Let lw, ld, lv and F be the log values of width, depth, velocity and discharge, respectively. Assuming n observations in each equation, one way to combine them is

$$\begin{pmatrix} lw_1 \\ ld_1 \\ lv_1 \\ \vdots \\ lw_n \\ ld_n \\ lv_n \end{pmatrix} = \begin{pmatrix} 1 & F_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & F_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & F_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & F_n & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & F_n & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & F_n \end{pmatrix} \begin{pmatrix} \alpha_w \\ \beta_w \\ \alpha_d \\ \beta_d \\ \alpha_v \\ \beta_v \end{pmatrix} + \begin{pmatrix} \epsilon_{w1} \\ \epsilon_{d1} \\ \epsilon_{v1} \\ \vdots \\ \epsilon_{wn} \\ \epsilon_{dn} \\ \epsilon_{vn} \end{pmatrix} \tag{2.22}$$

It should be pointed out that the set-up in equation(2.22) is specific to system (1.3). SUR method allows responses to depend on different sets of independent variables.

The parameter estimates are subject to $\boldsymbol{r} = \boldsymbol{R\beta}$, if the constraints were true, where $\boldsymbol{r} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $\boldsymbol{R} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$.

If each equation of system (1.3) satisfies the assumptions of a standard linear model, the disturbance vector $\boldsymbol{\epsilon}$ of equation(2.22) has the following characteristics:

1. Elements of the vector have zero mean and different variances. It is reasonable to assume that generally speaking, σ_w^2, σ_d^2 and σ_v^2 are unequal.

2. Disturbances at different times in the same equation, e.g., ϵ_{w1} and ϵ_{w2}, are uncorrelated.

3. Disturbances at different times across equations, e.g., ϵ_{w1} and ϵ_{d2}, are uncorrelated.

4. If there is reason to believe that correlation exists between disturbances across equations at the same time, e.g., between ϵ_{w1} and ϵ_{d1}, their covariance, denoted as σ_{wd}, is the so-called contemporaneous covariance.

Therefore, the covariance matrix of $\boldsymbol{\epsilon}$, in the presence of contemporaneous correlation, is block-diagonal with n diagonal sub-matrices:

$$\boldsymbol{\nu} = \begin{bmatrix} \boldsymbol{\Sigma} & 0 & \dots & 0 \\ 0 & \boldsymbol{\Sigma} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \boldsymbol{\Sigma} \end{bmatrix}, \text{ where } \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_w^2 & \sigma_{wd} & \sigma_{wv} \\ \sigma_{wd} & \sigma_d^2 & \sigma_{dv} \\ \sigma_{wv} & \sigma_{dv} & \sigma_v^2 \end{bmatrix}$$

The diagonal elements of $\boldsymbol{\Sigma}$ are independent of time, since the disturbances are assumed homoscedastic within each equation. The contemporaneous covariances are time-invariant as well, which implies that the errors come from a multivariate distribution with zero mean and a constant covariance matrix. Other elements of the matrix are 0, since they correspond to covariance of disturbances at different time points.

In order to make $\boldsymbol{\Sigma}$ nonsingular, we assume no linear dependence between any random pair of contemporaneous errors. The inverse of $\boldsymbol{\nu}$, $\boldsymbol{\nu}^{-1}$, is again a block-diagonal matrix with diagonal element $\boldsymbol{\Sigma}^{-1}$.

A computationally more convenient procedure can be set up as follows:

$$\begin{pmatrix} \boldsymbol{y_1} \\ \boldsymbol{y_2} \\ \vdots \\ \boldsymbol{y_L} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X_1} & 0 & \dots & 0 \\ 0 & \boldsymbol{X_2} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \boldsymbol{X_L} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta_1} \\ \boldsymbol{\beta_2} \\ \vdots \\ \boldsymbol{\beta_L} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon_1} \\ \boldsymbol{\epsilon_2} \\ \vdots \\ \boldsymbol{\epsilon_L} \end{pmatrix} \tag{2.23}$$

when the system has L equations to be estimated simultaneously. Each equation takes the form $\boldsymbol{y_i} = \boldsymbol{X_i}\boldsymbol{\beta_i} + \boldsymbol{\epsilon_i}, i = 1, \dots, L$, where $\boldsymbol{y_i}$ is a vector of the dependent variable of the i_{th} equation, $\boldsymbol{X_i}$ the corresponding values of explanatory variables, $\boldsymbol{\beta_i}$ the parameter vector and $\boldsymbol{\epsilon_i}$ the disturbance. The combined equations can then still be expressed as the basic form $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$.

For the hydraulic geometry problem, $L = 3$. The new set-up does not lead to different parameter estimates, however, the covariance matrix of $\boldsymbol{\epsilon}$ is no longer block-diagonal:

$$\boldsymbol{\nu} = \begin{bmatrix} cov(\boldsymbol{\epsilon_1}, \boldsymbol{\epsilon_1}) & cov(\boldsymbol{\epsilon_1}, \boldsymbol{\epsilon_2}) & cov(\boldsymbol{\epsilon_1}, \boldsymbol{\epsilon_3}) \\ cov(\boldsymbol{\epsilon_1}, \boldsymbol{\epsilon_2}) & cov(\boldsymbol{\epsilon_2}, \boldsymbol{\epsilon_2}) & cov(\boldsymbol{\epsilon_2}, \boldsymbol{\epsilon_3}) \\ cov(\boldsymbol{\epsilon_1}, \boldsymbol{\epsilon_3}) & cov(\boldsymbol{\epsilon_2}, \boldsymbol{\epsilon_3}) & cov(\boldsymbol{\epsilon_3}, \boldsymbol{\epsilon_3}) \end{bmatrix} = \begin{bmatrix} \sigma_{11}\boldsymbol{I} & \sigma_{12}\boldsymbol{I} & \sigma_{13}\boldsymbol{I} \\ \sigma_{12}\boldsymbol{I} & \sigma_{22}\boldsymbol{I} & \sigma_{23}\boldsymbol{I} \\ \sigma_{13}\boldsymbol{I} & \sigma_{23}\boldsymbol{I} & \sigma_{33}\boldsymbol{I} \end{bmatrix}$$

$cov(\boldsymbol{\epsilon_i}, \boldsymbol{\epsilon_j})$ is an $n \times n$ diagonal matrix with σ_{ij} down the diagonal. The off-diagonal elements are zero, because they correspond to disturbances at different time points.

It is convenient to express $\boldsymbol{\nu}$ in terms of Kronecker Product. The Kronecker Product of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, denoted as $\boldsymbol{A} \otimes \boldsymbol{B}$, is obtained by multiplying each element of $\boldsymbol{A}$ by the entire matrix $\boldsymbol{B}$. Using this notation,

$$\boldsymbol{\nu} = \boldsymbol{\Sigma} \otimes \boldsymbol{I}, \text{ where } \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix}$$

and its inverse, $\boldsymbol{\nu}^{-1}$, can be expressed as $\boldsymbol{\Sigma^{-1}} \otimes \boldsymbol{I}$. Clearly, $\boldsymbol{\Sigma}$ is the same as defined earlier.

The computation of SUR estimate is identical to that of $\hat{\boldsymbol{\beta}}_{\boldsymbol{gls}}$, as defined in equation(2.8), except that $\boldsymbol{V}$ is replaced by $\boldsymbol{\nu}$ and $\sigma^2 = 1$. Linear constraints testing under SUR is also the same as under single-equation GLS, except that $n - K$ is replaced by $Ln - \sum K_i$, where K_i is the number of regressors in the i^{th} equation.

The SUR method proceeds as follows: starting with an initial OLS regression, it uses OLS residuals to estimate the cross-equation covariance matrix $\boldsymbol{\Sigma}$. The estimator of $\boldsymbol{\Sigma}$ is computed as:

$$\boldsymbol{S} = \frac{1}{n} \begin{bmatrix} \sum_{j=1}^{n} \epsilon_{1j}\epsilon'_{1j} & \sum_{j=1}^{n} \epsilon_{1j}\epsilon'_{2j} & \sum_{j=1}^{n} \epsilon_{1j}\epsilon'_{3j} \\ \sum_{j=1}^{n} \epsilon_{1j}\epsilon'_{2j} & \sum_{j=1}^{n} \epsilon_{2j}\epsilon'_{2j} & \sum_{j=1}^{n} \epsilon_{2j}\epsilon'_{3j} \\ \sum_{j=1}^{n} \epsilon_{1j}\epsilon'_{3j} & \sum_{j=1}^{n} \epsilon_{2j}\epsilon'_{3j} & \sum_{j=1}^{n} \epsilon_{3j}\epsilon'_{3j} \end{bmatrix} \tag{2.24}$$

What will happen to SUR estimate in the absence of contemporaneous correlation? The short answer is that it leads to no gain over OLS, since the covariance matrix of $\boldsymbol{\epsilon}$ is then diagonal. The proof is simple:

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{\boldsymbol{gls}} &= [\boldsymbol{X}'\boldsymbol{\nu}^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{\nu}^{-1}\boldsymbol{y} \\
&= \left[\begin{pmatrix} \boldsymbol{X_1'} & 0 & \dots & 0 \\ 0 & \boldsymbol{X_2'} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \boldsymbol{X_L'} \end{pmatrix} \begin{pmatrix} \sigma^{11}\boldsymbol{I} & 0 & \dots & 0 \\ 0 & \sigma^{22}\boldsymbol{I} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \sigma^{LL}\boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{X_1} & 0 & \dots & 0 \\ 0 & \boldsymbol{X_2} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \boldsymbol{X_L} \end{pmatrix}\right]^{-1} \\
&\cdot \begin{pmatrix} \boldsymbol{X_1} & 0 & \dots & 0 \\ 0 & \boldsymbol{X_2} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \boldsymbol{X_L} \end{pmatrix} \begin{pmatrix} \sigma^{11}\boldsymbol{I} & 0 & \dots & 0 \\ 0 & \sigma^{22}\boldsymbol{I} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \sigma^{LL}\boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{y_1} \\ \boldsymbol{y_2} \\ \vdots \\ \boldsymbol{y_L} \end{pmatrix} \\
&= \begin{bmatrix} (\boldsymbol{X_1'X_1})^{-1}\boldsymbol{X_1'y_1} \\ (\boldsymbol{X_2'X_2})^{-1}\boldsymbol{X_2'y_2} \\ \vdots \\ (\boldsymbol{X_L'X_L})^{-1}\boldsymbol{X_L'y_L} \end{bmatrix} \\
&= (\boldsymbol{X'X})^{-1}\boldsymbol{X'y}
\end{aligned}$$

which is exactly the OLS estimate, and can be obtained from estimating the equations separately.

Another case where SUR has no advantage over OLS is when responses depend on the same set of explanatory variables. To see this, we formulate our results in terms of the

Kronecker Product, and apply one of its properties

$$\boldsymbol{A} \otimes (\boldsymbol{B} \otimes \boldsymbol{C}) = (\boldsymbol{A} \otimes \boldsymbol{B}) \otimes \boldsymbol{C}$$

The variance of SUR estimate is

$$V(\hat{\boldsymbol{\beta}}_{\boldsymbol{gls}}) = (\boldsymbol{X}'\boldsymbol{\nu}^{-1}\boldsymbol{X})^{-1} = [(\boldsymbol{I} \otimes \boldsymbol{x})'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I})(\boldsymbol{I} \otimes \boldsymbol{x})]^{-1} = \boldsymbol{\Sigma} \otimes (\boldsymbol{x}'\boldsymbol{x})^{-1}$$

The SUR estimate can then be expressed as

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\boldsymbol{gls}} &= (\boldsymbol{X}'\boldsymbol{\nu}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\nu}^{-1}\boldsymbol{y} \\ &= [\boldsymbol{\Sigma} \otimes (\boldsymbol{x}'\boldsymbol{x})^{-1}](\boldsymbol{I} \otimes \boldsymbol{x})'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I})\boldsymbol{y} = [\boldsymbol{I} \otimes (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}']\boldsymbol{y} \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}\end{aligned}$$

Therefore, when explanatory variables are identical, SUR leads to no gain over simple OLS.

Since OLS is as capable and adequate as SUR in the case of identical regressors, why would we consider applying SUR to the hydraulic geometry problem? The answer is that the restricted estimates, as defined in equation(2.5) and (2.13), are likely different under OLS and SUR, and so are the respective conclusions associated with linear hypothesis testing.

2.5 Inverse Prediction

It is of great interest to regulatory agencies to determine stream flow requirements for aquatic habitat protection. Such requirements might be obtained through inverse prediction from our hydraulic geometry model, if the model is well-calibrated to field observations. The advantage of inverse prediction is that it does not require stream flow records, and the stream flow requirements thus developed can be applied to hydrologically disturbed drainage basins and at gaged or ungaged sites.

For a regression equation

$$y = \beta_0 + \beta_1 X$$

that satisfies all assumptions of a standard linear regression model, the predicted mean value of y for a given X_0 is

$$\hat{y_0} = \hat{\beta}_0 + \hat{\beta}_1 X_0$$

and its 95% confidence interval is given by

$$\hat{y_0} \pm t_{n-K,1-\frac{1}{2}\alpha} \cdot s\{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum(X_i - \bar{X})^2}\}^{1/2}$$

After a regression line is fitted, the inverse estimate of X corresponding to a specified true mean value of y, say y_0, is given by

$$\hat{X}_0 = (y_0 - \hat{\beta}_0)/\hat{\beta}_1 \tag{2.25}$$

The lower limit of X, denoted X_L, can be obtained from solving the equation

$$y_0 = y_{x_L} - t_{n-K,1-\frac{1}{2}\alpha} \cdot s\{\frac{1}{n} + \frac{(X_L - \bar{X})^2}{\sum(X_i - \bar{X})^2}\}^{1/2}$$

where $y_{x_L} = \beta_0 + \beta_1 X_L$. Similarly, The upper limit of X, denoted X_U, is the solution to

$$y_0 = y_{x_U} + t_{n-K,1-\frac{1}{2}\alpha} \cdot s\{\frac{1}{n} + \frac{(X_U - \bar{X})^2}{\sum(X_i - \bar{X})^2}\}^{1/2}$$

where $y_{x_U} = \beta_0 + \beta_1 X_U$. After some manipulation, X_U and X_L are found to be

$$\hat{X}_0 + \frac{(\hat{X}_0 - \bar{X})g \pm t_{n-K,1-\frac{1}{2}\alpha} \cdot s/\hat{\beta}_1\{(X_0 - \bar{X})^2/\sum(X_i - \bar{X})^2 + (1-g)/n\}^{1/2}}{1-g} \quad (2.26)$$

where

$$g = \frac{t^2_{n-K,1-\frac{1}{2}\alpha} \cdot s^2}{\hat{\beta}_1^{\,2} \sum(X_i - \bar{X})^2} = \frac{t^2_{n-K,1-\frac{1}{2}\alpha}}{\frac{\hat{\beta}_1^{\,2}}{s^2/\sum(X_i - \bar{X})^2}} = (\frac{t_{n-K,1-\frac{1}{2}\alpha}}{\text{t-statistic of } \hat{\beta}_1})^2 \quad (2.27)$$

Clearly, the more precise is $\hat{\beta}_1$, the smaller is g. If the regression is not well-determined, then the confidence limits of X_0, as specified by equation(2.27), may not be real. Or the roots may be real but they fall on the same side of the regression line. The former case is illustrated with figure 2.1 on the next page. When true mean values of y are in the range of (-17, -12), the inverse confidence limits of X are not real.

Such inverse prediction method has not been developed for multivariate responses with contemporaneous correlations. Therefore, for regression coefficients obtained from SUR, equation(2.27) is used to determine the inverse confidence intervals. The degree of freedom does not change, i.e., it remains to be $n - K$, where n is the sample size in a single regression, and K is the number of explanatory variables in that regression.

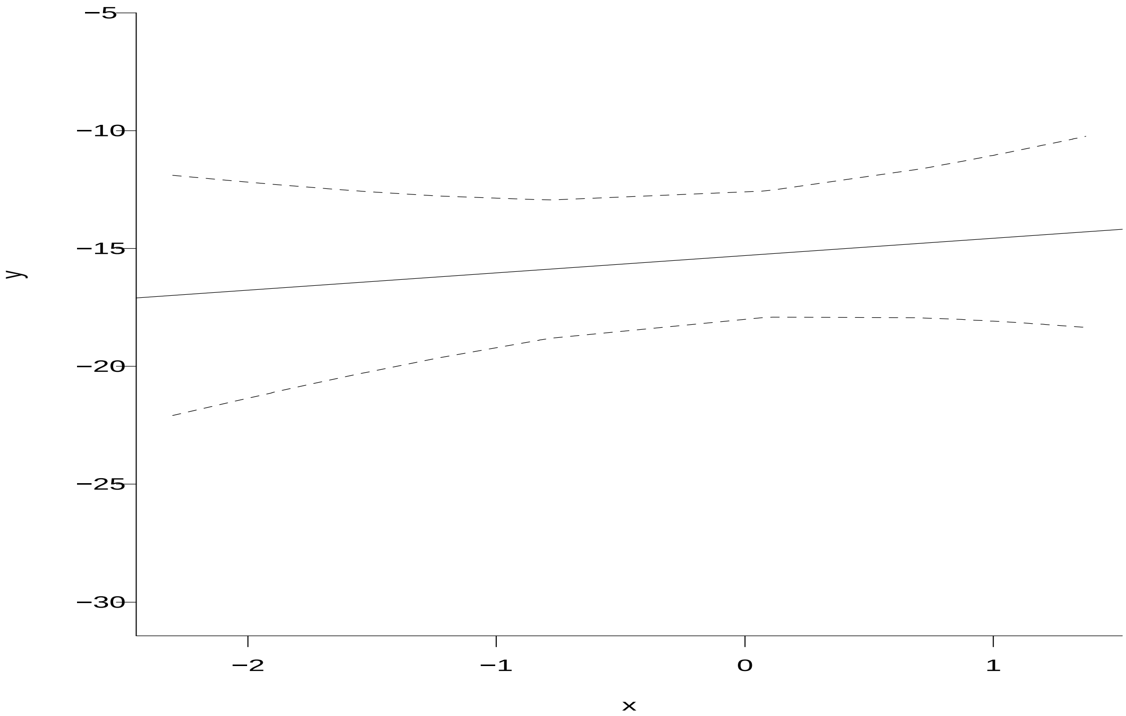

Figure 2.1: Inverse Regression Peculiarity: Complex Roots. The dashed lines represent the confidence intervals for the mean response at real value of x

Chapter 3

Numerical Examples

Table 3.1 provides raw data on cross-section 1 of Poole Creek located northeast of Pemberton in British Columbia, Canada.

Time	Total Discharge (m^3/s)	Wetted Width (m)	Mean Depth (m)	Mean Velocity (m/s)
July 2, 2002	2.76	11.60	0.26	0.92
Aug.14, 2002	2.00	8.50	0.30	0.79
Oct.31, 2002	0.13	4.90	0.13	0.21
Nov.21, 2002	0.29	6.70	0.15	0.29
Feb.11, 2003	0.17	5.30	0.12	0.26
May 11, 2003	0.44	7.00	0.16	0.39
May 27, 2003	2.61	9.25	0.34	0.82
May 29, 2003	3.74	10.10	0.37	1.01
May 31, 2003	3.35	10.55	0.35	0.92

Table 3.1: Time Series Data on Poole Creek Cross-section 1

Data collection was led by Scott Babakaiff at the BC Ministry of Water, Land and Air Protection. Once a cross-section was selected, they manually measured discharge using a rod-based current meter. Width was measured from bank to bank using a field tape measure. Multiple measurements were taken of depth with the same measuring device to calculate the mean depth, as depth varied depending on which part of the stream was being measured. To obtain velocity data, they used a flowmeter and moved it around, as velocity also varied depending on which part of the stream was being measured. Mean velocity was calculated as the average reading of multiple measurements that encompassed all parts of the cross-section.

The 27 observations are combined as follows:

$$\begin{pmatrix} log(11.60) \\ \vdots \\ log(10.55) \\ log(0.26) \\ \vdots \\ log(0.35) \\ log(0.92) \\ \vdots \\ log(0.92) \end{pmatrix} = \begin{pmatrix} 1 & log(2.76) & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & log(3.35) & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & log(2.76) & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & log(3.35) & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & log(2.76) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 1 & log(3.35) \end{pmatrix} \begin{pmatrix} \alpha_w \\ \beta_w \\ \alpha_d \\ \beta_d \\ \alpha_v \\ \beta_v \end{pmatrix} + \begin{pmatrix} \epsilon_{w1} \\ \vdots \\ \epsilon_{w9} \\ \epsilon_{d1} \\ \vdots \\ \epsilon_{d9} \\ \epsilon_{v1} \\ \vdots \\ \epsilon_{v9} \end{pmatrix}$$

In terms of the basic form $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$, $\boldsymbol{y}$ is a 27×1 vector taking the log values of width, depth and velocity; $\boldsymbol{X}$ is of order 27×6 with 3 block diagonal matrix

$$\boldsymbol{x} = \begin{pmatrix} 1 & log(2.76) \\ \vdots & \vdots \\ 1 & log(3.35) \end{pmatrix}$$

$\boldsymbol{\beta}$ is a 6×1 vector of coefficients, and $\boldsymbol{\epsilon}$ is a 27×1 disturbance vector.

The two linear constraints take the form

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{pmatrix} \alpha_w \\ \beta_w \\ \alpha_d \\ \beta_d \\ \alpha_v \\ \beta_v \end{pmatrix} \tag{3.1}$$

3.1 Using Matrix Methods Directly

The theory in Chapter 2 was used directly by programming in matrix manipulation in IML (SAS version 8.1, SAS Institute Inc., Cary, NC). IML code is attached in the Appendix, and estimates are presented in Table 3.2.

Parameter	Unrestricted	OLS_R	SUR_R
α_w	2.0813 (0.0287)	2.0795 (0.0230)	2.0898 (0.0282)
α_d	-1.4795 (0.0341)	-1.4820 (0.0240)	-1.4989 (0.0322)
α_v	-0.5969 (0.0176)	-0.5976 (0.0164)	-0.5909 (0.0173)
β_w	0.2157 (0.0221)	0.2169 (0.0177)	0.2098 (0.0218)
β_d	0.3229 (0.0264)	0.3246 (0.0185)	0.3363 (0.0248)
β_v	0.4581 (0.0136)	0.4585 (0.0127)	0.4539 (0.0134)
$\sum \alpha$	0.0049	0	0
$\sum \beta$	0.9966	1	1

Table 3.2: Estimates under OLS and SUR (standard errors in parentheses) for Poole Creek Cross Section 1

Since the unrestricted estimates are the same for OLS and SUR, the estimate of

$\boldsymbol{r} - \boldsymbol{R\beta}$ is always

$$\boldsymbol{r} - \boldsymbol{R}\hat{\boldsymbol{\beta}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 0.0049 \\ 0.9966 \end{bmatrix} = \begin{bmatrix} -0.0049 \\ 0.0034 \end{bmatrix}$$

The estimate of $\boldsymbol{\Sigma}$ under OLS is

$$\boldsymbol{S} = \begin{bmatrix} 0.0074 & 0 & 0 \\ 0 & 0.0104 & 0 \\ 0 & 0 & 0.0028 \end{bmatrix}$$

It is worth mentioning that $\boldsymbol{S}$ has been adjusted to account for the loss of degrees of freedom, because of the small sample size for the data in Table(3.1). Each equation has 9 observations and 2 explanatory variables, thus correcting for the degrees of freedom has an appreciable effect on the coefficient estimate. The adjustment becomes tricky when K_i is different for all i. If we were to divide different elements of $\boldsymbol{\epsilon}'\boldsymbol{M}\boldsymbol{\epsilon}$ by different $n - K_i$, the resulting $\boldsymbol{S}$ is not necessarily positive definite (Davidson and Mackinnon, 2004). In large samples, however, such correction is unnecessary.

Based on equation(2.21), the observed test statistic on the joint constraints is

$$\frac{3 \times 9 - 2 - 2 - 2}{2} \times \frac{0.0198}{21} = 0.0099$$

The null hypothesis is unlikely to be rejected as its p-value is 0.99.

Under SUR procedure, based on equation(2.24), $\boldsymbol{\Sigma}$ is replaced by

$$\boldsymbol{S} = \begin{bmatrix} 0.0074 & -0.0074 & -0.0001 \\ -0.0074 & 0.0104 & -0.0028 \\ -0.0001 & -0.0028 & 0.0028 \end{bmatrix}$$

An indication of the presence of correlation is that the correlation coefficient between width and depth residuals obtained from separate OLS regressions is -0.84 with p-value

0.005. The observed test statistic associated with SUR on the constraints is

$$\frac{3 \times 9 - 2 - 2 - 2}{2} \times \frac{5.5349}{21} = 2.7674$$

The data seems consistent with the null hypothesis, because the test statistic has a p-value equal to 0.086.

The restricted estimates are very close to the unrestricted, and we fail to reject the restrictions. However, the restricted SUR estimates are less efficient than the restricted OLS. Theoretically, SUR estimates will always be at least as efficient as OLS. But the asymptotic efficiency of SUR estimators may not carry over to small samples, because of the variability introduced by the estimated $\mathbf{\Sigma}$ (Greene, 2002).

Since all variables in the LLM model are expressed in logarithms, the estimated parameters can be interpreted independently of the unit of measurements. The slope coefficients in the LLM are rates of percentage change. Based on the restricted SUR model, about 45% of the increase in flow is accommodated by an increase in velocity, 34% by an increase in depth and 21% by an increase in width.

Figure 3.1 to 3.3 plot the three equations with the unrestricted fit, the restricted fit by OLS and the restricted fit by SUR superimposed.

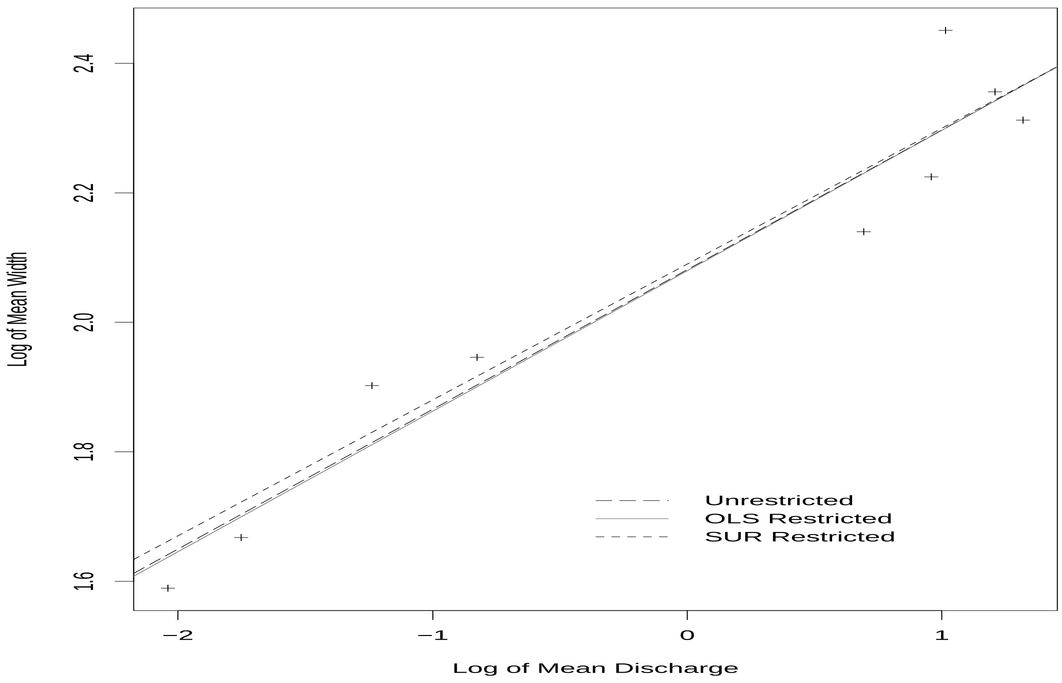

Figure 3.1: The Width Model fitted With OLS and SUR

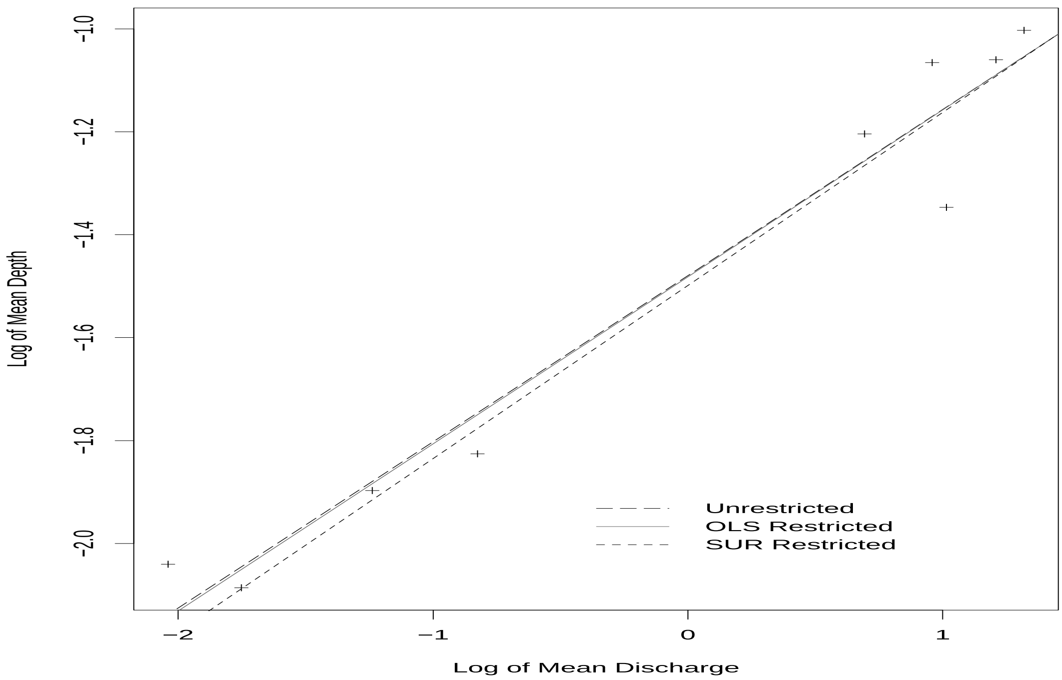

Figure 3.2: The Depth Model fitted With OLS and SUR

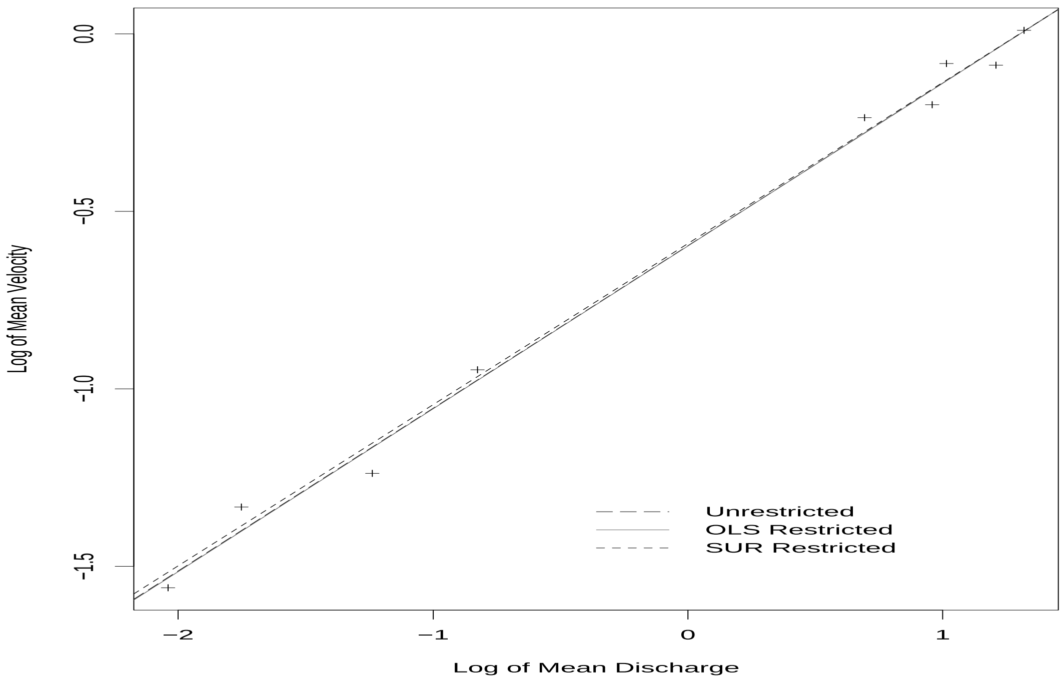

Figure 3.3: The Velocity Model fitted With OLS and SUR

3.2 Using the SYSLIN Procedure

SAS also offers a standardized procedure called SYSLIN which estimates parameters in an interdependent system of linear regression equations.

The estimation methods available to the SYSLIN procedure include OLS, SUR, Full Information Maximum Likelihood(FIML), etc. The STEST statement allows the user to test linear hypothesis on parameters in different models. The SRESTRICT statement imposes linear hypotheses on parameters in different models. The STEST statement and the SRESRTICT statement are not supported for the FIML estimation method.

3.2.1 OLS Regression with Restrictions

When no method of estimation is specified, PROC SYSLIN performs OLS regression:

```
DATA POOLE1;
    SET POOLE; IF CROSSSECTION = 1;
    FLOW = LOG(DISCHARGE);
    LW = LOG(WIDTH); LD = LOG(DEPTH); LV = LOG(VELOCITY);
RUN;
PROC SYSLIN DATA=POOLE1;
    MODEL LW = FLOW;
    MODEL LD = FLOW;
    MODEL LV = FLOW;
    STEST LW.INTERCEPT + LD.INTERCEPT + LV.INTERCEPT,
              LW.FLOW + LD.FLOW + LV.FLOW = 1;
```

```
RUN;

PROC SYSLIN DATA=POOLE1;
    MODEL LW = FLOW;
    MODEL LD = FLOW;
    MODEL LV = FLOW;
    SRESTRICT LW.INTERCEPT + LD.INTERCEPT + LV.INTERCEPT,
              LW.FLOW + LD.FLOW + LV.FLOW = 1;
RUN;
```

PROC SYSLIN with the STEST statement outputs unrestricted OLS estimates and tests the restrictions with an F test. When the STEST statement is in use with the SRESTRICT statement, the test statistic computed is conditional on the restrictions imposed. In our case, they cannot be used together, because then the F test will always be self-fulfilling.

PROC SYSLIN with the SRESTRICT statement outputs restricted OLS estimates. The Parameter Estimate table for the restricted model contains two additional rows for the restrictions specified by the SRESTRICT statement.

```
The SYSLIN Procedure
Ordinary Least Squares Estimation
```

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
RESTRICT	-1	2.237925	20.89819	0.11	0.9242
RESTRICT	-1	-2.62484	27.05259	-0.10	0.9313

The parameter estimates for the restrictions are values of the Lagrange multipliers used to impose the restrictions, and the test is a variant of the Lagrange Multiplier(LM) test. To see why it is called the LM test, consider a Lagrangian function:

$$\boldsymbol{l}(\boldsymbol{\beta}) - \boldsymbol{\lambda}'(\boldsymbol{R\beta} - \gamma)$$

where $\boldsymbol{l}(\boldsymbol{\beta})$ is the log-likelihood function. The solution $(\hat{\boldsymbol{\beta}}^{*}, \hat{\boldsymbol{\lambda}})$ satisfies the first order conditions

$$\frac{\partial \boldsymbol{l}(\hat{\boldsymbol{\beta}}^{*})}{\partial \boldsymbol{\beta}} - \boldsymbol{R}'\hat{\boldsymbol{\lambda}} = 0$$
$$\boldsymbol{R}\hat{\boldsymbol{\beta}}^{*} - \gamma = 0$$

$\hat{\boldsymbol{\lambda}}$ equals to $\hat{\boldsymbol{\phi}}$ in equation(2.12) scaled by a factor $\frac{1}{2}$. This comes straight from the log-likelihood function:

$$\boldsymbol{l}(\boldsymbol{\beta}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{\Sigma} \otimes \boldsymbol{I}|) - \frac{1}{2}[(\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{\Sigma} \otimes \boldsymbol{I})^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})] \tag{3.2}$$

Therefore,

$$\hat{\boldsymbol{\lambda}} = \frac{1}{2}\hat{\boldsymbol{\phi}} = -(\boldsymbol{RCR}')^{-1}(\gamma - \boldsymbol{R}\hat{\boldsymbol{\beta}}) \tag{3.3}$$

where $\hat{\boldsymbol{\beta}}$ is the unrestricted estimate.

The covariance matrix of $\hat{\boldsymbol{\lambda}}$ equals $(\boldsymbol{RCR}')^{-1}$. To see this, write $\hat{\boldsymbol{\lambda}}$ as a function of the $\boldsymbol{\epsilon}$'s:

$$\hat{\boldsymbol{\lambda}} = -(\boldsymbol{RCR}')^{-1}\boldsymbol{RCX}'(\boldsymbol{\Sigma} \otimes \boldsymbol{I})^{-1}\boldsymbol{\epsilon}$$

Proof of its variance follows directly after the fact that the expectation of $\hat{\boldsymbol{\lambda}}$ is 0. It then follows that the LM test takes the form

$$\hat{\boldsymbol{\lambda}}'[(\boldsymbol{RCR}')^{-1}]^{-1}\hat{\boldsymbol{\lambda}} = \hat{\boldsymbol{\lambda}}'\boldsymbol{RCR}'\hat{\boldsymbol{\lambda}} \tag{3.4}$$

Alternatively, the LM test can be expressed as

$$\frac{\partial \boldsymbol{l}(\hat{\boldsymbol{\beta}}^{*})}{\partial \boldsymbol{\beta}'}\mathbf{I}(\hat{\boldsymbol{\beta}}^{*})^{-1}\frac{\partial \boldsymbol{l}(\hat{\boldsymbol{\beta}}^{*})}{\partial \boldsymbol{\beta}} \tag{3.5}$$

where $\mathbf{I}(\boldsymbol{\beta})$ is the information matrix of $\boldsymbol{\beta}$. The underlying motivation of the test is that if the restrictions were true, the restricted estimate should be close to the unrestricted, and $\partial \boldsymbol{l}(\hat{\boldsymbol{\beta}}^*)/\partial \boldsymbol{\beta}$ should be close to 0. The LM test is in essence Rao's Score test and is distributed as χ^2_q where q is the number of constraints.

Since the likelihood functions of the restricted and the unrestricted estimates are the same, their information matrices must be the same. Therefore $\mathbf{I}(\hat{\boldsymbol{\beta}}^*)^{-\mathbf{1}} = \mathbf{I}(\hat{\boldsymbol{\beta}})^{-\mathbf{1}} = V(\hat{\boldsymbol{\beta}}) = \boldsymbol{C}$. The two forms of the LM test are further connected by the fact that $\partial \boldsymbol{l}(\hat{\boldsymbol{\beta}}^*)/\partial \boldsymbol{\beta} = \boldsymbol{R}'\hat{\boldsymbol{\lambda}}$, which comes directly from the first order conditions.

The SYSLIN procedure with SRESTRICT statement uses the first form to separately test the restrictions. The computation is the standard t test where the numerator is an individual Lagrange multiplier estimate and the denominator is the square root of the corresponding diagonal element of $(\boldsymbol{RCR}')^{-\mathbf{1}}$.

The p-value reported for a restriction is computed from a beta distribution rather than t-distribution, because the numerator and the denominator of the t-ratio for an estimated Lagrange multiplier are not independent (LaMotte, 1994). The p-values for restrictions seem highly insignificant, which imply that the data are consistent with the restrictions.

3.2.2 SUR with Restrictions

The SUR estimates are produced when the estimation method is specified as SUR.

```
PROC SYSLIN DATA=POOLE1 SUR;
    MODEL LW = FLOW;
    MODEL LD = FLOW;
    MODEL LV = FLOW;
    STEST LW.INTERCEPT + LD.INTERCEPT + LV.INTERCEPT,
              LW.FLOW + LD.FLOW + LV.FLOW = 1;
RUN;
PROC SYSLIN DATA=POOLE1 SUR;
    MODEL LW = FLOW;
    MODEL LD = FLOW;
    MODEL LV = FLOW;
    SRESTRICT LW.INTERCEPT + LD.INTERCEPT + LV.INTERCEPT,
              LW.FLOW + LD.FLOW + LV.FLOW = 1;
RUN;
```

PROC SYSLIN first produces unconstrained OLS results, because SUR requires OLS residuals to compute its covariance matrix. The output contains an estimate of the cross-model covariance $\mathbf{\Sigma}$, which has been presented in the previous section. The output also contains an estimate of the cross-model correlation of the disturbances, as an indication of the strength of contemporaneous correlation. The correlation between width and depth seems particularly strong, yet the correlations between the other pairs are either moderate or weak.

```
The SYSLIN Procedure
Seemingly Unrelated Regression Estimation
```

Cross Model Correlation

	lw	ld	lv
lw	1.00000	-0.84153	-0.02557
ld	-0.84153	1.00000	-0.51243
lv	-0.02557	-0.51243	1.00000

The Parameter Estimate table contains separate tests on the constraints. The constraints seem much weaker than under the OLS assumptions, although both p-values are moderately large.

```
The SYSLIN Procedure
Seemingly Unrelated Regression Estimation
```

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
RESTRICT	-1	626.5841	349.6839	1.79	0.0650
RESTRICT	-1	-734.914	452.6639	-1.62	0.1056

3.3 Results from Eight Other Cross Sections

OLS and SUR methods are applied to eight other cross sections on five creeks. Data was obtained from Scott Babakaiff at the BC Ministry of Water, Land and Air Protection. In all cross sections, at least one response pair exhibited significant contemporaneous correlation (Table 3.3). In particular, cross section 3 on the Poole Creek and cross section 1 on the Mashiter Creek had significant correlations among all responses. We would expect greater variance reduction from SUR in these two cross sections.

Cross Section	Width, Depth		Width, Velocity		Depth, Velocity	
Poole 1	-0.84	(0.005)	–		–	
Poole 2	–		-0.77	(0.01)	–	
Poole 3	-0.86	(0.001)	0.70	(0.03)	-0.96	(<0.0001)
Poole 4	–		-0.96	(0.0002)	–	
Stawamas 1	–		-0.996	(<0.0001)	–	
Mashiter 1	-0.96	(<0.0001)	0.91	(0.0006)	-0.99	(<0.0001)
Owl 1	–		-0.92	(0.001)	–	
Gold 2	-0.81	(0.008)	–		-0.80	(0.01)
Gold 3	–		–		-0.99	(<0.0001)

Table 3.3: Significant Contemporaneous Correlations among all Possible Pairs of Responses (p values in parentheses)

Table 3.4 summarizes hydraulic geometry estimates from OLS and SUR. SUR cannot be applied to cross sections 2 and 3 on the Gold Creek, because their contemporaneous covariance matrices of $\boldsymbol{\epsilon}$ are singular.

Table 3.4: Summary of Hydraulic Geometry Estimates
(standard errors in parentheses)

Cross Section	Parameter	Unrestricted	OLS_R	SUR_R
Poole 2	α_w	2.2612 (0.0286)	2.2608 (0.0226)	2.2662 (0.0264)
	α_d	-1.3435 (0.0199)	-1.3436 (0.0180)	-1.3420 (0.0196)
	α_v	-0.9167 (0.0312)	-0.9171 (0.0233)	-0.9242 (0.0266)
	β_w	0.1886 (0.0223)	0.1880 (0.0177)	0.1958 (0.0206)
	β_d	0.2591 (0.0155)	0.2589 (0.0140)	0.2612 (0.0153)
	β_v	0.5537 (0.0243)	0.5531 (0.0181)	0.5430 (0.0207)
	$\sum \alpha$	0.0010	0	0
	$\sum \beta$	1.0014	1	1
Poole 3	α_w	2.0297 (0.0098)	2.0296 (0.0095)	2.0329 (0.0087)
	α_d	-0.9547 (0.0367)	-0.9563 (0.0235)	-0.9576 (0.0364)
	α_v	-1.0723 (0.0290)	-1.0733 (0.0230)	-1.0753 (0.0287)
	β_w	0.0605 (0.0071)	0.0605 (0.0070)	0.0606 (0.0063)
	β_d	0.2910 (0.0268)	0.2910 (0.0171)	0.2910 (0.0266)
	β_v	0.6486 (0.0211)	0.6485 (0.0168)	0.6485 (0.0209)
	$\sum \alpha$	0.0027	0	0
	$\sum \beta$	1.0001	1	1
Poole 4	α_w	1.9744 (0.0060)	1.9745 (0.0059)	1.9734 (0.0058)
	α_d	-1.5088 (0.0219)	-1.5075 (0.0168)	-1.5176 (0.0188)
	α_v	-0.4687 (0.0255)	-0.4670 (0.0170)	-0.4559 (0.0195)
	β_w	0.1612 (0.0045)	0.1611 (0.0045)	0.1628 (0.0044)

continued on next page

continued from previous page

Cross Section	Parameter	Unrestricted	OLS_R	SUR_R
	β_d	0.4707 (0.0167)	0.4687 (0.0128)	0.4843 (0.0143)
	β_v	0.3728 (0.0194)	0.3702 (0.0129)	0.3530 (0.0148)
	$\sum \alpha$	-0.0031	0	0
	$\sum \beta$	1.0048	1	1
Stawamus 1	α_w	2.9962 (0.0095)	2.9962 (0.0095)	2.9923 (0.0081)
	α_d	-1.4757 (0.1011)	-1.4788 (0.0715)	-1.4806 (0.1009)
	α_v	-1.5142 (0.1007)	-1.5174 (0.0715)	-1.5118 (0.1006)
	β_w	0.0389 (0.0107)	0.0389 (0.0107)	0.0474 (0.0091)
	β_d	0.5592 (0.1136)	0.5661 (0.0804)	0.5700 (0.1134)
	β_v	0.3881 (0.1132)	0.3950 (0.0804)	0.3826 (0.1131)
	$\sum \alpha$	0.0063	0	0
	$\sum \beta$	0.9862	1	1
Mashiter 1	α_w	2.3430 (0.0248)	2.3431 (0.0245)	2.3404 (0.0235)
	α_d	-1.4126 (0.1138)	-1.4115 (0.0722)	-1.4074 (0.1126)
	α_v	-0.9323 (0.0900)	-0.9316 (0.0712)	-0.9330 (0.0900)
	β_w	0.0409 (0.0482)	0.0405 (0.0475)	0.0598 (0.0456)
	β_d	0.4716 (0.2208)	0.4637 (0.1400)	0.4343 (0.2186)
	β_v	0.5007 (0.1746)	0.4957 (0.1382)	0.5058 (0.1746)
	$\sum \alpha$	-0.0018	0	0
	$\sum \beta$	1.0132	1	1
Owl 1	α_w	2.3468 (0.0227)	2.3469 (0.0216)	2.3459 (0.0220)
	α_d	-1.3196 (0.0412)	-1.3192 (0.0343)	-1.3247 (0.0284)

continued on next page

continued from previous page

Cross Section	Parameter	Unrestricted	OLS_R	SUR_R
	α_v	-1.0284 (0.0573)	-1.0277 (0.0364)	-1.0212 (0.0385)
	β_w	0.1474 (0.0207)	0.1462 (0.0197)	0.1567 (0.0201)
	β_d	0.2565 (0.0376)	0.2525 (0.0313)	0.3090 (0.0259)
	β_v	0.6091 (0.0524)	0.6014 (0.0332)	0.5343 (0.0351)
	$\sum \alpha$	-0.0013	0	0
	$\sum \beta$	1.0129	1	1
Gold 2	α_w	3.3071 (0.0275)	3.3071 (0.0242)	–
	α_d	-1.5245 (0.0436)	-1.5245 (0.0288)	–
	α_v	-1.7826 (0.0267)	-1.7826 (0.0237)	–
	β_w	0.0717 (0.0204)	0.0717 (0.0180)	–
	β_d	0.3579 (0.0324)	0.3579 (0.0214)	–
	β_v	0.5703 (0.0199)	0.5703 (0.0176)	–
	$\sum \alpha$	0	0	–
	$\sum \beta$	1	1	–
Gold 3	α_w	3.2400 (0.0136)	3.2400 (0.0135)	–
	α_d	-1.7261 (0.0897)	-1.7261 (0.0661)	–
	α_v	-1.5140 (0.0969)	-1.5140 (0.0662)	–
	β_w	0.0269 (0.0107)	0.0269 (0.0107)	–
	β_d	0.5228 (0.0711)	0.5228 (0.0524)	–
	β_v	0.4503 (0.0769)	0.4503 (0.0525)	–
	$\sum \alpha$	0	0	–
	$\sum \beta$	1	1	–

Contrary to our expectations, restricted SUR didn't gain much efficiency over the unrestricted estimates for cross section 3 on Poole Creek, except in the width equation, although the slope coefficients in all equations are significant. Restricted OLS achieved greater variance reduction in the depth and velocity equations than restricted SUR. This also occurred in cross section 1 on the Mashiter Creek, but there, only the velocity equation is relatively well-determined. As Davidson and Mackinnon (2004) have pointed out, replacing $\mathbf{\Sigma}$ with an estimator based on OLS estimates inevitably degrades the finite-sample properties of the SUR estimator.

Amemiya (1985) proposes a solution to the singularity problem. Suppose that the rank of $\mathbf{\Sigma}$ is $T < Ln$ and $\mathbf{\Lambda}$ is a diagonal matrix consisting of the T positive characteristic roots. Then there exists an orthogonal matrix $\mathbf{H} = (\mathbf{H}_1, \mathbf{H}_2)$, where $\mathbf{H}_1$ is $Ln \times T$ and $\mathbf{H}_2$ is $Ln \times (Ln - T)$, such that $\mathbf{H}_1'\mathbf{\Sigma}\mathbf{H}_1 = \mathbf{\Lambda}$, $\mathbf{H}_1'\mathbf{\Sigma}\mathbf{H}_2 = 0$, and $\mathbf{H}_2'\mathbf{\Sigma}\mathbf{H}_2 = 0$. If we premultiply both sides of equation $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$ by $\mathbf{H}'$, the resulting equation can be partitioned into two parts:

$$\mathbf{H}_1'\boldsymbol{y} = \mathbf{H}_1'\boldsymbol{X\beta} + \mathbf{H}_1'\boldsymbol{\epsilon} \tag{3.6}$$

$$\mathbf{H}_2'\boldsymbol{y} = \mathbf{H}_2'\boldsymbol{X\beta} \tag{3.7}$$

$\mathbf{H}_2'\boldsymbol{\epsilon}$ is a zero vector because $\mathrm{E}(\mathbf{H}_2'\boldsymbol{\epsilon\epsilon}'\mathbf{H}_2) = \mathbf{H}_2'\mathbf{\Sigma}\mathbf{H}_2 = 0$. The best linear unbiased estimator of $\boldsymbol{\beta}$ is SUR applied to equation(3.6) subject to linear constraints in equation(3.7). Equivalently, it is OLS applied to

$$\mathbf{\Lambda}^{-1/2}\mathbf{H}_1'\boldsymbol{y} = \mathbf{\Lambda}^{-1/2}\mathbf{H}_1'\boldsymbol{X\beta} + \mathbf{\Lambda}^{-1/2}\mathbf{H}_1'\boldsymbol{\epsilon} \tag{3.8}$$

subject to the same constraints.

In terms of linear hypothesis testing, under OLS, the two constraints are always consistent with the data. However, with SUR, the unit-sum constraint does not hold for cross section 1 of the Owl Creek. The reason behind rejection is not clear, given that all hydraulic relations are fitted well with SUR. One possible explanation is the cross section's unique morphology. The Owl Creek is characterized by a mean elevation (1138m) well above the other four creeks, and a much steeper gradient (3.6%) in the study reach. Its cross section 1 is further featured with a low-elevation side bar. Table 3.5 reports p values of individual restrictions.

Cross Section	p value of $\lambda_{\sum \alpha}$		p value of $\lambda_{\sum \beta}$	
	under OLS	under SUR	under OLS	under SUR
Poole 1	0.9242	0.0650	0.9313	0.1056
Poole 2	0.9883	0.7551	0.9737	0.4771
Poole 3	0.9586	0.5099	0.9965	0.9568
Poole 4	0.9198	0.3639	0.8636	0.0829
Stawamas 1	0.9910	0.8445	0.9479	0.2181
Mashiter 1	0.9930	0.8181	0.9676	0.2631
Owl 1	0.9655	0.6586	0.8642	0.0271
Gold 2	1.0000	–	1.0000	–
Gold 3	1.0000	–	1.0000	–

Table 3.5: p values of Individual Restrictions under OLS and SUR (λ stands for Lagrange multiplier of an individual restriction)

3.4 Application of Inverse Prediction

Of the hydraulic exponents estimates, i.e., $\hat{\beta}_w$, $\hat{\beta}_d$ and $\hat{\beta}_v$, $\hat{\beta}_v$ is the only exponent significant in all nine cross-sections, and in six cross-sections it's the most significant exponent. In cross-section 3 of the Gold Creek, $\hat{\beta}_d$ has the smallest p value, and in cross-section 4 of the Poole Creek, such significance goes to $\hat{\beta}_w$. This is the consensus reached by OLS and SUR. In cross-section 1 of the Stawamus Creek, $\hat{\beta}_w$ is the most significant exponent from restricted SUR and $\hat{\beta}_d$ is the most significant exponent from restricted OLS.

In cross-section 1 of the Poole Creek, all regressions are well-fitted with restricted SUR, in that the t-statistics of $\hat{\beta}_w$, $\hat{\beta}_d$ and $\hat{\beta}_v$ are 9.6, 13.6 and 33.9, respectively. It is then interesting to see that for an inverse estimate of discharge at 2 m^3/s, how the 95% confidence limits obtained from the three equations would differ.

Figure 3.4 plotted the log of depth against the log of discharge, a straight line fitted with restricted SUR and 95% confidence interval of the true mean values of log depth. Based on equation(2.25), an inverse estimate of discharge at 2 m^3/s corresponds to a true mean depth of 0.282 meter.We draw a horizontal line at a height log(0.282). The intersections of the horizontal line and the confidence bands of y give us the inverse confidence interval of log(discharge). Solving for x_L and x_U in equation(2.26), the inverse CI of log(discharge) is (0.43, 1.00), and the corresponding inverse CI of discharge at a mean depth of 0.282 meter is (1.54 m^3/s, 2.72 m^3/s).

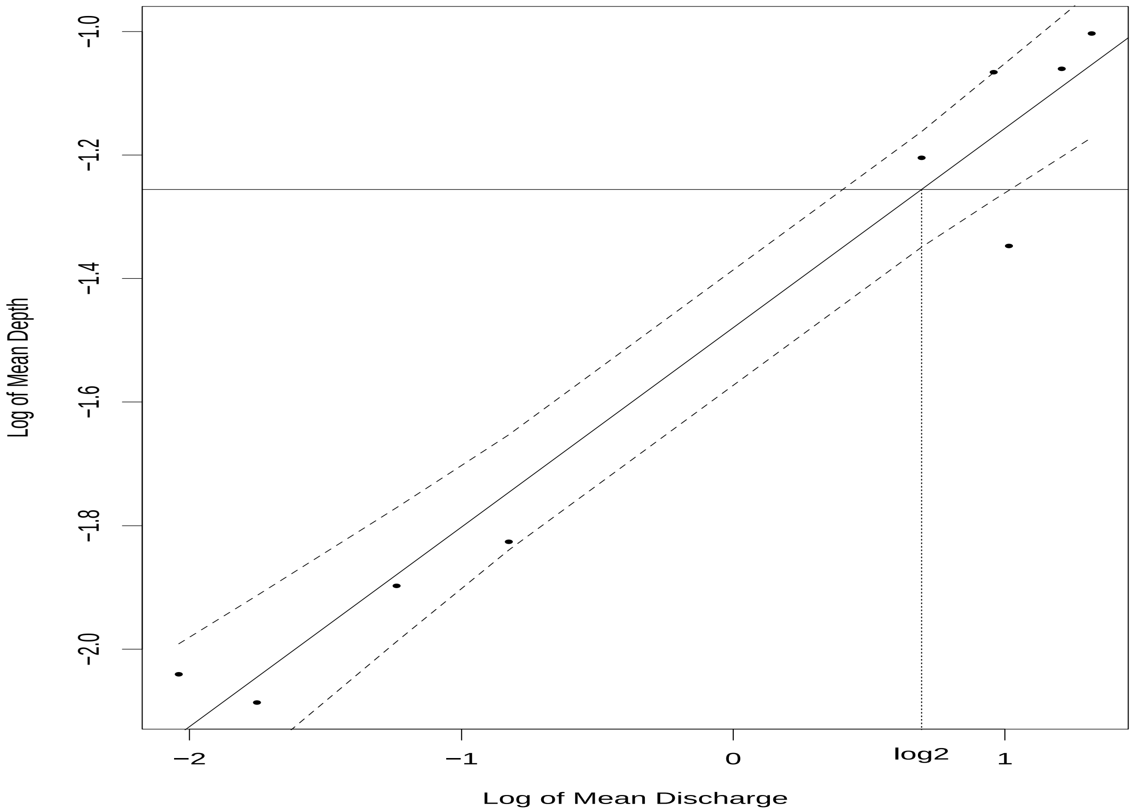

Figure 3.4: The fitted regression line (——) log(depth) = -1.4988 + 0.3363*log(discharge) and 95% confidence bands (- - -) of mean log(depth) – A restricted SUR model was used to fit the model.

When we repeat the same procedure with the width model, for an estimated discharge at 2 m^3/s, its log of CI is (0.35, 1.13) which yields an inverse 95% CI of (1.42 m^3/s, 3.10 m^3/s).

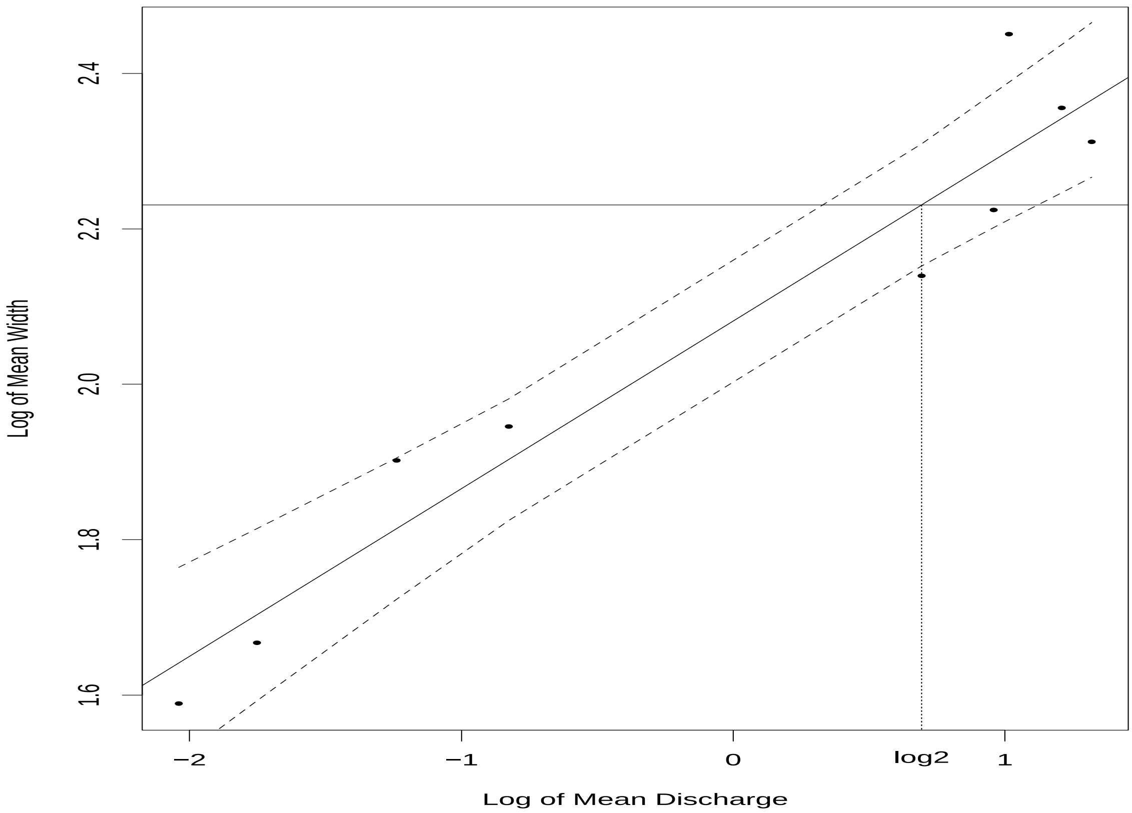

Figure 3.5: The fitted regression line (——) log(width) = 2.0898 + 0.2098*log(discharge) and 95% confidence bands (- - -) of mean log(width) – A restricted SUR model was used to fit the model.

Using the velocity model, for an estimated discharge at 2 m^3/s, its log of CI is (0.59, 0.80) which transforms to a CI of discharge equaling (1.80 m^3/s, 2.23 m^3/s).

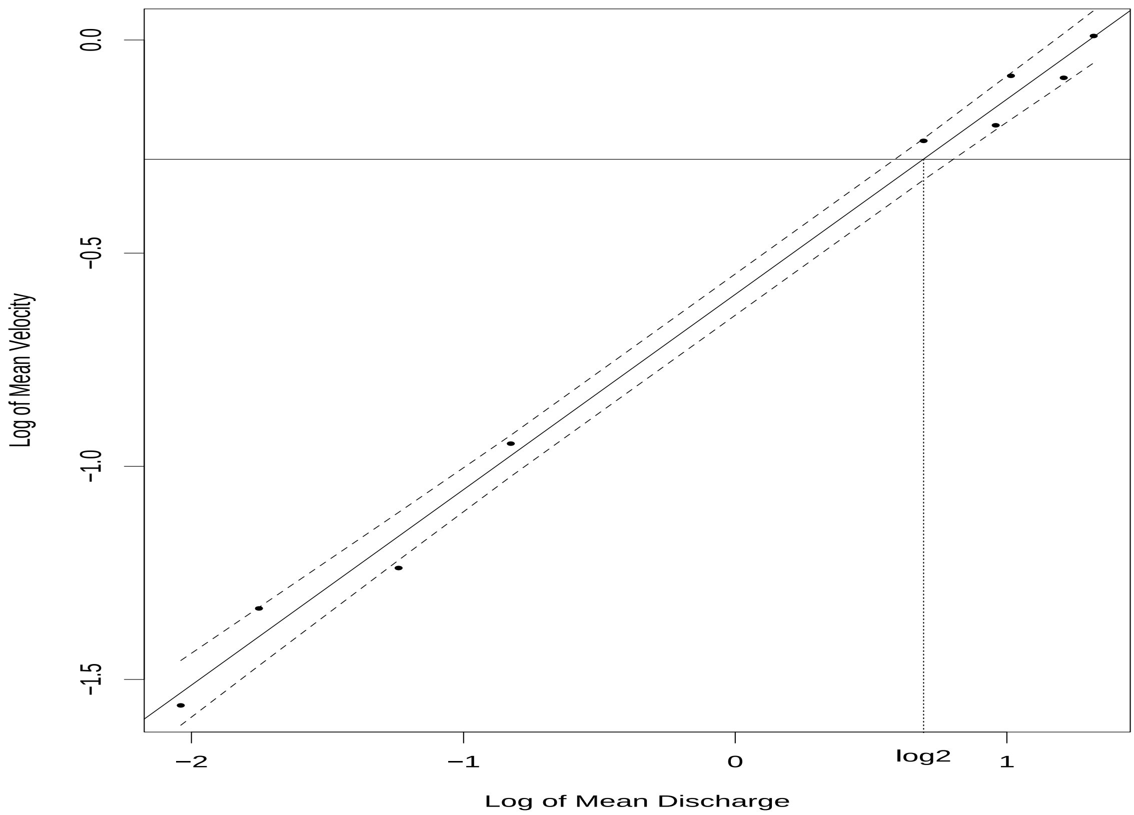

Figure 3.6: The fitted regression line (——) log(velocity) = -0.5909 + 0.4539*log(discharge) and 95% confidence bands (- - -) of mean log(velocity) – A restricted SUR model was used to fit the model.

The scatter of the dots around the fitted line is directly related to the width of the confidence band, and to width of the confidence interval for the inverse prediction. When width, depth and velocity data are all available and at equal cost, velocity seems a preferable choice under most circumstances for inverse prediction purpose. That being said, the t-statistic of $\hat{\beta}_v$ in the velocity model is 33.9, which is highly significant. Consequently, the predicted inverse CI of discharge from the velocity model seems pretty narrow.

Chapter 4

Summary

In this paper, we have showed how to perform joint-equation estimation and place linear constraints across equations. Traditional literature, using single-equation estimation method, have either ignored the constraints, or force the exponents to sum to unity by arbitrarily manipulating the estimates. By using a systematic approach called Seemingly Unrelated Regression(SUR), we are able to jointly estimate the relations and impose the constraints. Seemingly Unrelated Regression method was compared to Ordinary Least Squares.

The advantage of SUR is that it allows for correlations across equations, it is capable of estimating simultaneously the complete set of parameters in the system, and it makes it possible to test and impose cross-equation constraints. The major drawback of SUR, and other systematic approaches, is that the consequence of misspecification, such as leaving out important explanatory variables or including redundant variables in the model, is more severe under SUR than single-equation estimation technique. With SUR method, if one equation is misspecified, the poor estimates for that equation may contaminate estimates for other equations, whereas if equation-by-equation OLS is employed, only the estimates of that equation are affected.

It is well known that if the errors across equations are uncorrelated or if the sets of regressors are identical, the unrestricted SUR has no payoff against the unrestricted OLS. On the other hand, the greater the correlation of the error and the less the correlation of the regressors, the greater the efficiency gain (Greene, 2002).

Theoretically, SUR is at least as efficient as OLS. In finite samples however, estimating the cross-equation covariance from OLS residuals increases the sampling variation of the covariance, thus SUR may be less efficient. Zellner(1963) investigated a two-equation system for which the regressors of different equations are orthogonal. He stated that “only for contemporaneous correlation in the neighborhood of 0 and/or small values of n”, where n is the sample size after correcting for the degree of freedom, “that the single-equation OLS estimator is slightly superior”. He also conceded that when the orthogonality condition is not met, “there will be some reduction in the gain to be realized by application of the joint estimation procedure”. It is suggested that when the sample size is small and the cross-equation correlations are small, OLS is preferable to SUR.

We also briefly investigated the possibility of inverse prediction. When regression models are well-determined, i.e., when the explanatory variables are statistically significant, calibration of discharge from hydraulic geometry produces quite narrow inverse confidence intervals.

After relations of hydraulic geometry for a number of creeks are established, it will be interesting to see whether cross-sections estimates on the same creek can be combined, and if they are, whether some “regional” relations of hydraulic geometry can be developed from averaging the results.

Appendix: SAS IML Code

```
DATA POOLE1;
    SET POOLE; IF CROSSSECTION = 1;
    FLOW = LOG(DISCHARGE);
    LW = LOG(WIDTH); LD = LOG(DEPTH); LV = LOG(VELOCITY);
RUN;
PROC IML;
    USE POOLE1;
    READ ALL VAR {FLOW} INTO XVAR; READ ALL VAR {LW LD LV} INTO YVAR;
    y = YVAR[,1]//YVAR[,2]//YVAR[,3];
        * Design matrix of dependent variables in joint regressions;
    XX = (j(nrow(XVAR),1,1)||XVAR);
        * Design matrix of independent variables in a single regression;
    mat0 = {0 0, 0 0, 0 0, 0 0, 0 0, 0 0, 0 0, 0 0, 0 0};
    X = (XX||mat0||mat0)//(mat0||XX||mat0)//(mat0||mat0||XX);
        * Design matrix of independent variables in joint regressions;
    sr = {0, 1}; R = {1 0 1 0 1 0, 0 1 0 1 0 1};
        * Matrices to impose restrictions;
    DF_OLS = nrow(YVAR[,1]) - ncol(XX);
        * Degrees of freedom in a single regression;

    * Single-equation OLS regression;
    B_OLS_lw = (inv(t(XX)*XX)*t(XX)*YVAR[,1]); * Width regression;
    RES_OLS_lw = YVAR[,1] - XX*B_OLS_lw;
```

```
STD_B_OLS_lw = sqrt(vecdiag(inv(t(XX)*XX)*ssq(RES_OLS_lw)/DF_OLS));
B_OLS_ld = (inv(t(XX)*XX)*t(XX)*YVAR[,2]); * Depth regression;
RES_OLS_ld = YVAR[,2] - XX*B_OLS_ld;
STD_B_OLS_ld = sqrt(vecdiag(inv(t(XX)*XX)*ssq(RES_OLS_ld)/DF_OLS));
B_OLS_lv = (inv(t(XX)*XX)*t(XX)*YVAR[,3]); * Velocity regression;
RES_OLS_lv = YVAR[,3] - XX*B_OLS_lv;
STD_B_OLS_lv = sqrt(vecdiag(inv(t(XX)*XX)*ssq(RES_OLS_lv)/DF_OLS));
B_OLS = B_OLS_lw//B_OLS_ld//B_OLS_lv;
    * Combine single-equation estimates;
STD_B_OLS = STD_B_OLS_lw//STD_B_OLS_ld//STD_B_OLS_lv;
    * Combine single-equation standard errors;

* Joint-equation OLS regressions:
    parameter estimates, standard errors, t-stats and p-values;
B_OLS_2 = inv(t(X)*X)*t(X)*y;
RES_OLS = y - X*B_OLS;
STD_B_OLS_2 = sqrt(vecdiag(inv(t(X)*X)))#(sqrt((ssq(RES_OLS[1:9])
    //ssq(RES_OLS[1:9])//ssq(RES_OLS[10:18]) //ssq(RES_OLS[10:18])
    //ssq(RES_OLS[19:27])//ssq(RES_OLS[19:27]))/DF_OLS));
T_B_OLS = B_OLS#(1/STD_B_OLS);
PVAL_B_OLS = 1 - probF(T_B_OLS#T_B_OLS, 1, DF_OLS);

* Inverse of estimator of Sigma assuming no correlation.
    The matrix is diagonal;
INVDIAGS = inv(((ssq(RES_OLS[1:9])||0||0)//(0||ssq(RES_OLS[10:18])
    ||0)//(0||0||ssq(RES_OLS[19:27])))/DF_OLS)@I(9);
* Joint GLS regressions assuming no correlation;
B_OLS_3 = inv(t(X)*INVDIAGS*X)*t(X)*INVDIAGS*y;
STD_B_OLS_3 = sqrt(vecdiag(inv(t(X)*INVDIAGS*X)));

* F test on the joint restrictions under OLS regression;
```

```
F_TEST_OLS = t((sr-R*B_OLS))*inv(R*inv(t(X)*INVDIAGS*X)*t(R))
    *(sr-R*B_OLS)/(t(RES_OLS)*INVDIAGS*RES_OLS)*DF_OLS*3/2;
PVAL_F_OLS = 1-probF(F_TEST_OLS, 2, DF_OLS*3);

* Restricted OLS estimates;
B_OLS_R = B_OLS + inv(t(X)*INVDIAGS*X)*t(R)*inv(R*inv(t(X)*
    INVDIAGS*X)*t(R))*(sr-R*B_OLS);
STD_B_OLS_R = sqrt(vecdiag(inv(t(X)*INVDIAGS*X) - inv(t(X)*
    INVDIAGS*X)*t(R)*inv(R*inv(t(X)*INVDIAGS*X)*t(R))*R*inv(t(X)
    *INVDIAGS*X)));
T_B_OLS_R = B_OLS_R#(1/STD_B_OLS_R);
PVAL_B_OLS_R = 1 - probF(T_B_OLS_R#T_B_OLS_R, 1, DF_OLS);

* t-test on the individual restrictions under OLS;
LAMDA_OLS = -inv(R*inv(t(X)*INVDIAGS*X)*t(R))*(sr-R*B_OLS);
STD_LAMDA_OLS = sqrt(vecdiag(inv(R*inv(t(X)*INVDIAGS*X)*t(R))));
T_LAMDA_OLS = LAMDA_OLS#(1/STD_LAMDA_OLS);

PRINT B_OLS STD_B_OLS B_OLS_2 STD_B_OLS_2 B_OLS_3 STD_B_OLS_3;
PRINT T_B_OLS PVAL_B_OLS F_TEST_OLS PVAL_F_OLS;
PRINT B_OLS_R STD_B_OLS_R T_B_OLS_R PVAL_B_OLS_R LAMDA_OLS
    STD_LAMDA_OLS T_LAMDA_OLS;

* Inverse of the estimator of Sigma allowing for contemporaneous
    correlation. The matrix is NOT diagonal;
INVGLSS = inv(((ssq(RES_OLS[1:9])||sum(RES_OLS[1:9]#RES_OLS
    [10:18])||sum(RES_OLS[1:9]#RES_OLS[19:27]))//(sum(RES_OLS
    [1:9]#RES_OLS[10:18])||ssq(RES_OLS[10:18])||sum(RES_OLS
    [10:18]#RES_OLS[19:27]))//(sum(RES_OLS[1:9]#RES_OLS[19:27])
    ||sum(RES_OLS[10:18]#RES_OLS[19:27])||ssq(RES_OLS[19:27])))
    /DF_OLS)@I(9);
```

```
* Joint GLS regressions assuming contemporaneous correlation:
    parameter estimates, standard errors, t-stats and p-values;
B_GLS = inv(t(X)*INVGLSS*X)*t(X)*INVGLSS*y;
STD_B_GLS = sqrt(vecdiag(inv(t(X)*INVGLSS*X)));
T_B_GLS = B_GLS#(1/STD_B_GLS);
PVAL_B_GLS = 1 - probF(T_B_GLS#T_B_GLS, 1, DF_OLS);

* F test on the joint restrictions under Joint GLS regressions;
F_TEST_GLS = t((sr-R*B_GLS))*inv(R*inv(t(X)*INVGLSS*X)*t(R))*
    (sr-R*B_GLS)/(t(RES_OLS)*INVGLSS*RES_OLS)*DF_OLS*3/2;
PVAL_F_GLS = 1-probF(F_TEST_GLS, 2, DF_OLS*3);

* Restricted Joint GLS estimates;
B_GLS_R = B_GLS + inv(t(X)*INVGLSS*X)*t(R)*inv(R*inv(t(X)*
    INVGLSS*X)*t(R))*(sr-R*B_GLS);
STD_B_GLS_R = sqrt(vecdiag(inv(t(X)*INVGLSS*X) - inv(t(X)*
    INVGLSS*X)*t(R)*inv(R*inv(t(X)*INVGLSS*X)*t(R))*R*inv(t(X)*
    INVGLSS*X)));
T_B_GLS_R = B_GLS_R#(1/STD_B_GLS_R);
PVAL_B_GLS_R = 1 - probF(T_B_GLS_R#T_B_GLS_R, 1, DF_OLS);

* t-test on the individual restrictions under Joint GLS;
LAMDA_GLS = -inv(R*inv(t(X)*INVGLSS*X)*t(R))*(sr-R*B_GLS);
STD_LAMDA_GLS = sqrt(vecdiag(inv(R*inv(t(X)*INVGLSS*X)*t(R))));
T_LAMDA_GLS = LAMDA_GLS#(1/STD_LAMDA_GLS);

PRINT B_GLS STD_B_GLS T_B_GLS PVAL_B_GLS F_TEST_GLS PVAL_F_GLS;
PRINT B_GLS_R STD_B_GLS_R T_B_GLS_R PVAL_B_GLS_R LAMDA_GLS
    STD_LAMDA_GLS T_LAMDA_GLS;

* Inverse prediction of CI when discharge = 2;
```

```
X0 = log(2); t_value = tinv(0.975, 7);
Xbar = sum(XVAR)/9; Sxx = ssq(XVAR - Xbar);
s_Width = sqrt(ssq(RES_OLS[1:9])/DF_OLS);
s_Depth = sqrt(ssq(RES_OLS[10:18])/DF_OLS);
s_Velocity = sqrt(ssq(RES_OLS[19:27])/DF_OLS);
print X0 t_value Xbar Sxx s_Width s_Depth s_Velocity;

g_Width= (t_value/T_B_GLS_R[2])*(t_value/T_B_GLS_R[2]);
Xu_Width= X0 + ((X0 - Xbar)*g_Width - (t_value*s_Width/B_GLS_R[2])
    *sqrt((X0 - Xbar)*(X0 - Xbar)/Sxx + (1-g_Width)
    /nrow(YVAR[,1])))/(1-g_Width);
Xl_Width= X0 + ((X0 - Xbar)*g_Width - (t_value*s_Width/B_GLS_R[2])
    *sqrt((X0 - Xbar)*(X0 - Xbar)/Sxx + (1-g_Width)
    /nrow(YVAR[,1])))/(1-g_Width);
Xuw = exp(Xu_Width); Xlw = exp(Xl_Width);
print g_Width Xl_Width Xu_Width Xlw Xuw;

g_Depth= (t_value/T_B_GLS_R[4])*(t_value/T_B_GLS_R[4]);
Xu_Depth= X0 + ((X0 - Xbar)*g_Depth + (t_value*s_Depth/B_GLS_R[4])
    *sqrt((X0 - Xbar)*(X0 - Xbar)/Sxx + (1-g_Depth)
    /nrow(YVAR[,1])))/(1-g_Depth);
Xl_Depth= X0 + ((X0 - Xbar)*g_Depth - (t_value*s_Depth/B_GLS_R[4])
    *sqrt((X0 - Xbar)*(X0 - Xbar)/Sxx + (1-g_Depth)
    /nrow(YVAR[,1])))/(1-g_Depth);
Xud = exp(Xu_Depth); Xld = exp(Xl_Depth);
print g_Depth Xl_Depth Xu_Depth Xld Xud;

g_Velocity= (t_value/T_B_GLS_R[6])*(t_value/T_B_GLS_R[6]);
Xu_Velocity= X0 + ((X0 - Xbar)*g_Velocity + (t_value*s_Velocity
    /B_GLS_R[6])*sqrt((X0 - Xbar)*(X0 - Xbar)/Sxx + (1-g_Velocity)
    /nrow(YVAR[,1])))/(1-g_Width);
```

```
    Xl_Velocity= X0 + ((X0 - Xbar)*g_Velocity - (t_value*s_Velocity
        /B_GLS_R[6])*sqrt((X0 - Xbar)*(X0 - Xbar)/Sxx + (1-g_Velocity)
        /nrow(YVAR[,1])))/(1-g_Width);
    Xuv = exp(Xu_Velocity); Xlv = exp(Xl_Velocity);
    print g_Velocity Xl_Velocity Xu_Velocity Xlv Xuv;
QUIT;
```

Bibliography

[1] Amemiya, T. *Advanced Econometrics.* Harvard University Press, Cambridge, MA, 1985.

[2] Babakaiff, S. At-a-Station hydraulic geometry: useful for considering instream flows or simply a bit of outdated empirical science? Unpublished manuscript, 2004.

[3] Bates, B. A statistical log piecewise linear model of at-a-station hydraulic geometry. *Water Resources Research*, 26:109-118, 1990.

[4] Davidson, R. and Mackinnon, J. *Econometric Theory and Methods.* Oxford University Press, New York, 2004.

[5] Draper, N. and Smith, H. *Applied Regression Analysis.* John Wiley, New York, 1981.

[6] Greene, W. *Econometric Analysis*, 5th edn. Macmillian, New York, 2002.

[7] Grunfeld, Y. The determinants of corporate investment. Unpublished thesis, Chicago, 1958.

[8] Jowett, L. Hydraulic geometry of New Zealand rivers and its use as a preliminary method of habitat assessment. *Regulated Rivers: Research*, 14:451-466, 1998.

[9] Judge, G., Griffiths, W., Hill, R., Lutkepohl, H. and Lee, T. *The Theory and Practice of Econometrics*, 2nd edn. John Wiley, New York, 1985.

[10] Kennedy, P. *A Guide to Econometrics*, 4th edn. Harvard University Press, Cambridge, MA, 1998.

[11] LaMotte, L. A note on the role of independence in t statistics constructed from linear statistics in regression models. *The American Statistician*, 48:238-240, 1994.

[12] Leopold, L. and Maddock, T. The hydraulic geometry of stream channels and some physiographic implications. *US Geological Survey Professional Paper*, 252-357, 1953.

[13] Rhodes, D. Hydraulic geometry and channel roughness - a non-linear system. *American Journal of Science*, 273:877-896, 1973.

[14] Rhodes, D. The b-f-m diagram: graphical representation and interpretation of at-a-station hydraulic geometry. *American Journal of Science*, 277:73-96, 1977.

[15] Richards, K. The hydraulic geometry of stream channels and some physiographic implications. *American Journal of Science*, 273:877-896, 1973.

[16] Ridenour, S. and Giardino, J. The statistical study of hydraulic geometry: a new direction for compositional data analysis. *Mathematical Geology*, 23:349-366, 1991.

[17] SAS Institute Inc. *SAS/ETS User's Guide*, Version 8. SAS Institute Inc., Cary, NC, 2000.

[18] Srivastava, V. and Giles, D. *Seemingly Unrelated Regression Equation Models: Estimation and Inference*. Marcel Dekker, New York, 1987.

[19] Theil, H. *Principles of Econometrics*. John Wiley, New York, 1971.

[20] Zellner, A. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57:348-368, 1962.

[21] Zellner, A. Estimators for seemingly unrelated regression equations: some exact finite sample results. *Journal of the American Statistical Association*, 58:977-992, 1963.