IMPOSING OBSERVATION-VARYING EQUALITY CONSTRAINTS USING GENERALISED RESTRICTED LEAST SQUARES

Howard E. Doran
School of Economics
University of New England
Armidale Australia

Christopher J. O'Donnell
&
Alicia N. Rambaldi
School of Economics
University of Queensland
Queensland 4072 Australia

April 2003

Discussion Paper No 323

ISSN 1446-5523

IMPOSING OBSERVATION-VARYING EQUALITY CONSTRAINTS USING GENERALISED RESTRICTED LEAST SQUARES

Howard E. Doran, Christopher J. O'Donnell & Alicia N. Rambaldi

Abstract:

Linear equality restrictions derived from economic theory are frequently observation-varying. Except in special cases, Restricted Least Squares (RLS) cannot be used to impose such restrictions without either underconstraining or overconstraining the parameter space. We solve the problem by developing a new estimator that collapses to RLS in cases where the restrictions are observation-invariant. We derive some theoretical properties of our so-called Generalised Restricted Least Squares (GRLS) estimator, and conduct a simulation experiment involving the estimation of a constant returns to scale production function. We find that GRLS significantly outperforms RLS in both small and large samples.

1. INTRODUCTION

Parameters in econometric models are often subject to 'extraneous information' (Goldberger (1964)) arising from economic theory. For example, a log-linear production function with constant returns to scale has the condition (or constraint) that the sum of the input coefficients equals one. Econometrics textbooks furnish many similar examples, which almost invariably involve constraints that do not vary across observations. The constraints are a necessary and sufficient expression of the 'extraneous information', and estimation of the parameters is completely standard, using for example, Restricted Least Squares (RLS).

However, there are other models in which the economically implied constraints do vary across observations. We illustrate with two examples:

First, in a study on inflation, Clements and Izan (1987), postulated a model in which the inflation rate for commodity i at time t could be decomposed into a time effect α_t and a commodity effect β_i. In order to identify the model, a constraint of the form $\Sigma_i w_{it}\beta_i = 0$ (where w_{it} is a known number for all i and t) had to be imposed. The essential feature of this constraint on the β_i is that it changes with t.

Second, systems of demand equations should satisfy homogeneity, Engel and Slutsky conditions. If the system model is of the very convenient constant elasticity form, both the Engel and Slutsky constraints on the parameters involve the expenditure shares of the commodities, and these vary over time. Furthermore, the Slutsky conditions are non-linear in the expenditure shares. We will return to this point subsequently.

A fundamental property of observation-varying constraints is that it is impossible to satisfy them in the context of fixed-parameter models. There are always more constraints than parameters. This fact has given rise to three estimation approaches.

First, because of the non-standard nature of the constraints, they are simply ignored (see Beattie and Taylor, 1985, p. 119). This is far from satisfactory as it is often the constraints that encapsulate economic behavioural assumptions. The unconstrained model usually does little more than select relevant variables and combine them in a form suitable for easy estimation and interpretation. The resulting estimation will be inefficient, and will not usually satisfy underlying economic theory.

Second, the observation-varying constraints are replaced by a set of fixed constraints that are sufficient to satisfy the original set (but are not necessary). The estimation problem then becomes standard. An example of this approach is discussed in Section IV. The difficulties with this method are that in many cases it may not be possible to find a sufficient set of fixed-parameter constraints, and even when such a set is available, the parameter space is over-constrained, leading to biased estimates.

Third, the number of constraints is reduced by assuming they apply only at the arithmetic or geometric mean of the data. Examples of this approach are Clements and Izan (1987) and Selvanathan (1989). Estimation becomes standard, but again, the parameter space is over-constrained. An additional problem here is that if the constraints are non-linear in the data, the artificial constraints will not be consistent with the original set, leading to more bias.

In this paper we propose a computationally simple, regression based method which allows the parameters also to vary across observations. This opens the possibility of obtaining estimates which exactly satisfy the constraints at every data point, but which still use the data to select estimates that are optimal in some sense.

The paper is structured as follows. In Section II we outline our general approach to estimating a varying-parameter model subject to observation-varying constraints. Our approach involves substituting the constraints into the model before making an invariance assumption that allows the parameters to be identified. In Section III we show how the general theory of linear equations can be used to implement the approach, and we motivate and derive the theoretical properties of a least squares estimator. Because this estimator collapses to RLS under certain conditions, we refer to it as Generalised Restricted Least Squares (GRLS). In Section IV we describe and report the results of a Monte Carlo experiment designed to compare the performance of GRLS and two RLS estimators of the parameters of a constant returns to scale production function. We find that GRLS dominates these RLS estimators in terms of bias and within-sample predictive performance. The paper is concluded in Section V.

2. GENERAL METHODOLOGY FOR IMPOSING OBSERVATION-VARYING CONSTRAINTS

Consider a set of *observation-varying* constraints, linear in an unknown parameter β, and written in the usual notation as

$$\boldsymbol{R}_t\beta = \mathbf{r}_t\,,\ t = 1, 2, \ldots T, \tag{1}$$

where β is of dimension $K \times 1$, $\mathbf{R}_t$ is $J \times K$, $\mathbf{r}_t$ is $J \times 1$, the rank of $\boldsymbol{R}_t$ is $J < K$, and both $\boldsymbol{R}_t$ and $\mathbf{r}_t$ are non-stochastic and known for all t. In total, there are JT constraints which we are seeking to impose on K parameters. Unless some constraints are redundant, this is not generally possible. As discussed in the Introduction, the usual approaches to the problem have involved changing the constraints so that their number is reduced. Our approach is to increase the number of unknown parameters, by allowing the β to be observation-varying.

Thus, our starting point is the linear model

$$\mathbf{y}_t = \boldsymbol{X}_t\boldsymbol{\beta}_t + \mathbf{e}_t, t = 1, ..., T, \tag{2}$$

where $\mathbf{y}_t$ is the t-th observation on an $N \times 1$ vector of endogenous variables ($N \geq 1$), $\boldsymbol{X}_t$ is an associated $N \times K$ design matrix, $\boldsymbol{\beta}_t$ is a $K \times 1$ vector of unknown parameters, and $\mathbf{e}_t$ is an $N \times 1$ disturbance vector. Without loss of generality, we assume $E\{\mathbf{e}_t\} = \mathbf{0}_N$ and $E\{\mathbf{e}_t\ \mathbf{e}_t'\} = \sigma_e^2\boldsymbol{I}_N$ where $\mathbf{0}_N$ is an $N \times 1$ vector of zeros, σ_e^2 is an unknown scalar, and $\boldsymbol{I}_N$ is an $N \times N$ identity matrix. The most distinctive feature of the model is that the parameter vector $\boldsymbol{\beta}_t$ varies across observations. Moreover, the constraints now take the form

$$\boldsymbol{R}_t\boldsymbol{\beta}_t = \mathbf{r}_t. \tag{3}$$

Model (2) is quite common in econometrics and a survey can be found in Judge *et al* (1985, ch 19). Noteworthy examples are systematically varying parameter models, switching regressions, piecewise regression models and Hildreth-Houck models. State-space models could also be added to this list. As it stands (2) is not identified, and in all the examples cited, identification is achieved by placing some additional (across-observation) structure on the $\boldsymbol{\beta}_t$. Of course, the most common example is the general linear model

$$\mathbf{y}_t = \boldsymbol{X}_t\boldsymbol{\beta} + \mathbf{e}_t, \tag{4}$$

which is simply (2) with the identifying condition $\boldsymbol{\beta}_t = \boldsymbol{\beta}$. This invariance assumption is so widespread in econometrics that the fixed-parameter general linear model is often regarded as the starting point for the linear model-building process, instead of being regarded as a special case of (2).

Our econometric problem is to estimate the *KT* observation-varying parameters in the model (2) subject to the observation-varying constraints given by (3). Our simple approach involves substituting the constraints (3) into the model (2) to obtain an *unconstrained* model of the form

$$\mathbf{w}_t = \boldsymbol{Z}_t \boldsymbol{\gamma}_t + \mathbf{e}_t \tag{5}$$

where $\mathbf{w}_t$ and $\boldsymbol{Z}_t$ are known transformations of $\mathbf{y}_t$ and $\boldsymbol{X}_t$, and $\boldsymbol{\gamma}_t$ is a new $K \times 1$ vector of unknown parameters which have a known relationship to $\boldsymbol{\beta}_t$.

Model (5) is, of course, unidentified, and we overcome this problem by making the identifying assumption

$$\boldsymbol{\gamma}_t = \boldsymbol{\gamma} \tag{6}$$

where $\boldsymbol{\gamma}$ is a $K \times 1$ vector of fixed parameters. Except in special cases, the parameters of the resulting model can be estimated using standard econometric techniques such as Ordinary Least Squares (OLS). The final step in our approach is to obtain estimates of $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_T$ using the known relationship between $\boldsymbol{\beta}_t$ and $\boldsymbol{\gamma}_t = \boldsymbol{\gamma}$. Importantly, these estimates will exactly satisfy the constraints (3).

It is important to note that our method does not *demand* the invariance of γ_t. It is straightforward to estimate the observation-varying parameter model (5) using, for example, Flexible Least Squares (see Kalaba and Tesfatsion, 1989). In other cases it may be more appropriate to use a more general identifying assumption of the form

$$\gamma_t = \delta\mathbf{j} + \rho\gamma_{t-1} + \mathbf{u}_t,\ |\rho| < 1 \tag{7}$$

where $\mathbf{j}$ is a $K \times 1$ unit vector, δ and ρ are scalars, and $\mathbf{u}_t$ is a $K \times 1$ random vector. In such cases, estimation could be effected by using the Kalman filter. However, in the remainder of this paper we will assume (6).

A distinguishing feature of our approach is that the identifying assumption (6) is an observation-invariance assumption on γ_t, a parameter vector that is theoretically unconstrained (the constraints have been substituted out in the transformation from $\mathbf{y}_t$ to $\mathbf{w}_t$). This contrasts with the usual RLS approach where the identifying assumption is an observation-invariance assumption on $\boldsymbol{\beta}_t$, a parameter vector that must satisfy the theoretical constraints given by (3). As already mentioned, this is problematic because there is no general solution to the constraints given by (1).

The alternative approach we are suggesting in this paper circumvents these difficulties by simply delaying the usual parametric invariance assumption until after the observation-varying constraints have been substituted into the model. This delay means that our invariance assumption is made on parameters that are theoretically unconstrained. It is possible to substitute the constraints into the model in several ways, and in the following section we describe a particular method that leads to an estimator with a number of desirable statistical properties.

III. Generalised Restricted Least Squares (GRLS)

In this section we show how the general theory of linear equations is used to substitute the constraints (3) into the model (2) to obtain the observation-varying parameter model (5). Under the identifying assumption (6), the parameters of this model can be estimated using standard techniques such as OLS. In this section we discuss one such OLS-based estimator and derive some of its more important properties.

Consider the constraints (3) where the rank of $\boldsymbol{R}_t$ is $J \leq K$. The general solution to (3) is (eg. Graybill, 1969, p.142):

$$\beta_t = \boldsymbol{R}_t^{+}\mathbf{r}_t + \boldsymbol{H}_t\gamma_t \tag{8}$$

where γ_t is an arbitrary $K \times 1$ vector, $\boldsymbol{R}_t^{+}$ is the (unique) Moore-Penrose generalised inverse of $\boldsymbol{R}_t$, and $\boldsymbol{H}_t \equiv \boldsymbol{I}_K - \boldsymbol{R}_t^{+}\boldsymbol{R}_t$ is a symmetric idempotent $K \times K$ matrix. A vector β_t will exactly satisfy the constraints (3) *if and only if* it has the form of (8). The estimation problem is to choose γ_t in some optimal way, and to this end we turn our attention to the information contained in the model (2).

In light of the result (8), the model (2) can be written in the form of (5) where $\mathbf{w}_t \equiv \mathbf{y}_t - \boldsymbol{X}_t\boldsymbol{R}_t^{+}\mathbf{r}_t$ is $N \times 1$ and $\boldsymbol{Z}_t \equiv \boldsymbol{X}_t\boldsymbol{H}_t$ is $N \times K$. Clearly, $\mathbf{w}_t$ and $\boldsymbol{Z}_t$ are observed for all t. As it stands, however, γ_t in the model (5) is not identified, so we make the invariance assumption (6) to obtain

$$\mathbf{w}_t = \boldsymbol{Z}_t\,\gamma + \mathbf{e}_t. \tag{9}$$

This is a standard fixed-parameter general linear model. Thus, if the matrix $\boldsymbol{Z} = (\boldsymbol{Z}_1', ..., \boldsymbol{Z}_T')'$ is of full column rank[1], an estimate of γ, denoted $\mathbf{g}$, can be obtained using any conventional econometric technique. Estimates of β_t can then be recovered from (8) as

$$\mathbf{b}_t = \boldsymbol{R}_t^{+}\mathbf{r}_t + \boldsymbol{H}_t\mathbf{g}. \tag{10}$$

In the remainder of this section we consider the OLS estimator of γ in the model (9), and we derive some properties of the associated estimator of β_t in the model (2).

Under our earlier assumptions on the error vector $\mathbf{e}_t$, the best linear unbiased estimator of γ in (9) is simply the OLS estimator

$$\mathbf{g} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\mathbf{w} \tag{11}$$

where $\mathbf{w} = (\mathbf{w}_1', ..., \mathbf{w}_T')'$. Note that equation (10) expresses $\mathbf{b}_t$ as a deterministic linear function of $\mathbf{g}$. Thus, if the parametric invariance assumption $\gamma_t = \gamma$ holds, the estimator $\mathbf{b}_t$ defined by (10) and (11) is the best linear unbiased estimator of β_t in (2) and (3).

The estimator defined by (10) and (11) has a number of other important properties. Proofs of the following three propositions are provided in the Appendix.

P.1 Let $\boldsymbol{R}_t = \boldsymbol{R}$ and $\mathbf{r}_t = \mathbf{r}$ for all t. Then the estimator defined by (10) and (11) is identical to the RLS estimator $\mathbf{b}_{RLS} = \mathbf{b} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\mathbf{b} - \mathbf{r})$ where $\boldsymbol{X} = (\boldsymbol{X}_1', ..., \boldsymbol{X}_T')'$, $\mathbf{y} = (\mathbf{y}_1', ..., \mathbf{y}_T')'$ and $\mathbf{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathbf{y}$.

---

[1] We are unaware of any result that establishes the full column rank of $\boldsymbol{Z}$, but in every empirical example we have seen, this property holds. If $\boldsymbol{Z}$ is not of full column rank, a unique estimator of γ can only be obtained by introducing more information into the estimation process.

P.2 Let $\boldsymbol{R}_t = [\boldsymbol{R}_{1t}, \mathbf{0}_{J,K_2}]$ where $\boldsymbol{R}_{1t}$ is $J \times K_1$ ($J < K_1$) of rank J and $K_2 \equiv K - K_1$. Let β_t and $\mathbf{b}_t$ be partitioned conformably as $\beta_t = [\beta_{1t}', \beta_{2t}']'$ and $\mathbf{b}_t = [\mathbf{b}_{1t}', \mathbf{b}_{2t}']'$. Then $\mathbf{b}_{2t}$ is always observation-invariant.

P.3 The estimator defined by (10) and (11) is invariant with respect to a re-ordering of the regressors.

Proposition P.1 says that our estimator collapses to conventional RLS if the constraints are observation-invariant. For this reason we refer to our estimator as Generalised Restricted Least Squares (GRLS). It is no surprise that the (best linear unbiased) GRLS estimator and the conventional RLS estimator are identical when the constraints are observation-invariant – in this case the assumption $\beta_t = \beta$ becomes feasible and the (unique) RLS estimator is known to be best linear unbiased if the constraints are true.

Proposition P.2 says that if any element of β_t is unconstrained by (3) (ie. if the constraints provide no information concerning the evolution of an element of β_t) then the corresponding element of $\mathbf{b}_t$ defined by (10) and (11) will be observation-invariant.

Proposition P.3 distinguishes the GRLS estimator from other estimators which use different methods to substitute the constraints (3) into the model (2). These other estimators invariably involve partitioning the vector β_t into observation-varying and observation-invariant subsets. The problem with these estimators is that the partitioning is totally arbitrary, implying the parameter estimates are not invariant to an arbitrary re-ordering of the regressors. Such an estimator has been used by O'Donnell, Shumway and Ball (1999).

IV. Monte Carlo Experiment

In this section we describe a simple experiment designed to investigate the properties of GRLS and RLS estimators of a constant returns to scale (CRS) production function. We have chosen this model for its familiarity, and because it provides an example of the 'sufficient conditions' mentioned in Section I.

The aims of the experiment are to compare the performance of the GRLS estimator proposed in this paper with two commonly used fixed parameter restricted estimators, namely, an estimator based on sufficient conditions and one obtained by replacing the variables in the restrictions by their sample means.

Our experiment is designed to examine relative performance in three cases, namely when

(i) The identifying assumption (6) holds (ie. $\gamma_t = \gamma$) and sufficient conditions of the form (1) are almost satisfied at every observation (ie. when the usual assumption $\beta_t = \beta$ is nearly feasible).

(ii) The identifying assumption (6) holds, but the sufficient conditions are not satisfied.

(iii) Neither the identifying assumption, nor the sufficient conditions hold.

We consider a translog production function defined over output y_t and inputs x_{1t} and x_{2t}:

$$\ln(y_t) = \beta_{0t} + \sum_{i=1}^{2} \beta_{it} \ln(x_{it}) + .5 \sum_{i=1}^{2} \sum_{j=1}^{2} \beta_{ijt} \ln(x_{it}) \ln(x_{jt}) + e_t, \qquad t = 1, ..., T, \tag{12}$$

where β_{0t}, β_{1t}, β_{2t}, β_{11t}, β_{12t}, β_{21t} and β_{22t} are parameters, $\beta_{12t} = \beta_{21t}$, and the e_t are iid disturbance terms with zero means and constant variance, σ_e^2. If the function exhibits CRS then, by Euler's Theorem, the parameters must satisfy

$$\sum_{i=1}^{2}\beta_{it} + \sum_{i=1}^{2}\sum_{j=1}^{2}\beta_{ijt}\ln(x_{jt}) = 1. \tag{13}$$

This condition is both necessary and sufficient for CRS, and implies observation-varying constraints on the parameters.

Note that (12) can be written in the form of (2), with $N = 1$, $K = 6$, $\boldsymbol{X}_t = [1 \quad \ln(x_{1t}) \quad \ln(x_{2t}) \quad .5\ln(x_{1t})^2 \quad \ln(x_{1t})\ln(x_{2t}) \quad .5\ln(x_{1t})^2]$ and $\boldsymbol{\beta}_t = (\beta_{0t}, \beta_{1t}, \beta_{2t}, \beta_{11t}, \beta_{12t}, \beta_{22t})'$. Similarly, (13) can be written in the form of (3) with $J = 1$, $\boldsymbol{R}_t = [0 \quad 1 \quad 1 \quad \ln(x_{1t}) \quad \ln(x_{1t}x_{2t}) \quad \ln(x_{2t})]$ and $\mathbf{r}_t = 1$. Finally, the result (8) means the model can also be compactly written in the form of (5).

Our Monte Carlo experiment is conducted by generating data according to (12) and (13) or, equivalently, according to (5). We begin by setting $x_{1t} = t$ and generating x_{2t} from an $N(\mu_x, \sigma_x^2)$ distribution. To limit the sample space of our experiment, we draw only one set of x_{2t} values, using $\mu_x = 500$ and $\sigma_x^2 = 1000$, and keep these values fixed across treatments and replications. To generate data on y_t we allow γ_t in (5) to evolve according to the general dynamic process (7), namely, $\gamma_t = \delta\mathbf{j} + \rho\gamma_{t-1} + \mathbf{u}_t$, $|\rho|<1$. This process includes, as special cases, most of the identifying assumptions found in the econometrics literature, and is sufficiently flexible to allow the performance comparisons outlined above.

The role of ρ in (7) is two-fold: in the case where $\sigma_u^2 = 0$, ρ has the effect of increasing the mean of γ_t; in the case where $\sigma_u^2 \neq 0$, ρ also has the effect of changing the variance of $\boldsymbol{\beta}_t$. In order to confine ρ to this role, and to rule out disequilibrium effects, we set γ_0 to its equilibrium value, $\delta\mathbf{j}/(1-\rho)$.

The control parameters in our experiment are δ, ρ, σ_u, σ_e and T. When $\sigma_u = 0$, the identifying assumption (6) holds exactly. Otherwise, we have chosen values of ρ, δ and σ_u so that the coefficient of variation CV = $(\sigma_u\sqrt{1-\rho})/(\delta\sqrt{1+\rho})$ is small, implying that (6) is at least a reasonable approximation to reality.

For each set of control parameter values of interest, we draw $\mathbf{u}_t$ from the $N(\mathbf{0}_6, \sigma_u^2\boldsymbol{I}_6)$ distribution, generate γ_t using (7), draw e_t from the $N(0, \sigma_e^2)$ distribution, generate y_t using (5), and estimate the parameter vector $\boldsymbol{\beta}_t$ using GRLS and two RLS estimators. For each set of control parameter values, these steps are replicated $N = 1000$ times.

In each replication of the experiment we estimate the model using the GRLS estimator given by (10) and (11). We also consider two RLS estimators, both of which are obtained under the parametric invariance assumption $\boldsymbol{\beta}_t = \boldsymbol{\beta}$ which, in the case of our translog production function, means

$$\beta_{0t} = \beta_0,\ \beta_{1t} = \beta_1,\ \beta_{2t} = \beta_2,\ \beta_{11t} = \beta_{11},\ \beta_{12t} = \beta_{12} \text{ and } \beta_{22t} = \beta_{22}. \tag{14}$$

Under this invariance constraint the CRS constraint (13) becomes

$$\sum_{i=1}^{2}\beta_i + \sum_{i=1}^{2}\sum_{j=1}^{2}\beta_{ij}\ln(x_{jt}) = 1 \tag{15}$$

which will be satisfied if

$$\sum_{i=1}^{2} \beta_i = 1 \text{ and } \sum_{i=1}^{2} \beta_{ij} = 0 \text{ for } j = 1, 2. \tag{16}$$

When econometricians set out to impose CRS on a single-output two-input translog production function, they typically impose the constraints (16) on the model given by (12) and (14). Such constraints are completely arbitrary and have no theoretical support. The corresponding RLS estimator, which we call RLS1, clearly overconstrains the parameter space, since the constraints (14) and (16) are *sufficient but not necessary* for the CRS constraint (13) to hold. In our Monte Carlo experiment we also consider an alternative RLS estimator (called RLS2) which imposes the constraint (15) at the arithmetic means of x_{1t} and x_{2t}. This RLS estimator underconstrains the parameter space insofar as the parameters are only required to satisfy the constraints (15) at a single point, which is unlikely to be a sample point.

The results of our Monte Carlo experiment are reported in Tables 1 and 2 for sample sizes $T = 50$ and $T = 400$. In the top section of these tables we report our assumed values of δ, ρ, σ_u, σ_e, CV and associated elements of

$$E\{\beta_t\} = \boldsymbol{R}_t^{+}\mathbf{r}_t + \delta \boldsymbol{H}_t \mathbf{j}/(1\text{-}\rho) \tag{17}$$

for $t = 1$ and $t = T$. This mean vector can be used to roughly assess the validity of the sufficient conditions given by (16), while the difference between $E\{\beta_1\}$ and $E\{\beta_T\}$ is a crude measure of the degree to which the parameters are observation-invariant. Note that (17) is constant across replications, even when $\sigma_u^2 \neq 0$. The top section of Tables 1 and 2 also reports

$$R\text{-squared} = (1/N) \sum_{n=1}^{N} \mathrm{Corr}[\ln(y_t), \mathrm{E}\{\ln(y_t)\}]^2 \tag{18}$$

which is a measure of the average proportion of the variation in $\ln(y_t)$ which is systematic. In the remaining parts of Tables 1 and 2 we report three standard measures of estimator performance, namely bias, Root Mean Square Error (RMSE) and mean *R*-squared.

In Tables 1 and 2 we focus on a subset of the elements of the representative vectors β_1 and β_T, partly to conserve space, and partly because our data generating process guarantees that $\mathrm{E}\{\beta_{0t}\} = \mathrm{E}\{\beta_{11,1}\}$, $\mathrm{E}\{\beta_{12,1}\} = \mathrm{E}\{\beta_{22,1}\}$ and $\mathrm{E}\{\beta_{1t}\} = \mathrm{E}\{\beta_{2t}\}$ for all t, so presenting information on all of these parameters is unnecessarily repetitive. We also focus on sets of control parameters which allow us to explore the robustness of our three estimators. In the remainder of this section we present a rationale for our chosen sets of control parameters, and we interpret the associated simulation results.

In column *A* of Table 1 we have set $\rho = \sigma_u^2 = 0$ to ensure γ_t is constant, $\sigma_e^2 = 0.25$ to ensure that approximately 80% of the variation in $\ln(y_t)$ is systematic, and $\delta = 0.65$ to ensure that the sufficient conditions given by (16) are close to being met for almost every t. The RLS1 and GRLS estimators can be expected to perform reasonably well with these settings, and this is evidenced in the lower sections of Table 1 where we have used asterisks to identify the smallest bias, lowest RMSE and highest *R*-squared statistics. The GRLS estimator appears to be least biased, while RLS1 appears to have lowest RMSE. The GRLS estimator is able to explain a marginally greater proportion of the variation in $\ln(y_t)$ than either of the

Table 1. – Monte Carlo Results for $T = 50$.

		A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
<u>True Values</u>										
ρ		0.00	0.20	0.80	0.00	0.20	0.80	0.80	0.00	0.80
σ_e		0.25	0.25	0.25	10.50	10.50	10.50	0.08	0.08	0.08
σ_u		0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.08	0.08
δ		0.65	0.65	0.65	50.00	50.00	50.00	0.65	50.00	50.00
CV		0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.002	0.0005
$E\{\beta_{0,1}\} = E\{\beta_{11,1}\} = E\{\beta_{0,T}\}$		0.65	0.81	3.25	50.00	62.50	250.00	3.25	50.00	250.00
$E\{\beta_{1,1}\} = E\{\beta_{2,1}\}$		0.48	0.59	2.25	34.21	42.75	170.90	2.25	34.21	170.90
$E\{\beta_{12,1}\} = E\{\beta_{22,1}\}$		0.01	-0.02	-0.47	-9.02	-11.31	-45.60	-0.47	-9.02	-45.60
$E\{\beta_{1,T}\} = E\{\beta_{2,T}\}$		0.53	0.67	2.63	40.26	50.33	201.27	2.63	40.26	201.27
$E\{\beta_{11,T}\}$		0.20	0.24	0.82	11.91	14.87	59.36	0.82	11.91	59.36
$E\{\beta_{12,T}\}$		-0.22	-0.30	-1.46	-23.65	-29.58	-118.58	-1.46	-23.65	-118.58
$E\{\beta_{22,T}\}$		0.23	0.28	0.98	14.44	18.05	72.06	0.98	14.44	72.06
R-squared		0.80	0.84	0.98	0.82	0.88	0.99	0.96	1.00†	1.00†
<u>Bias</u>										
RLS1	$\beta_{0,1}$	-0.10	-0.56	-7.41	-138.81	-174.03	-700.98*	-7.40	-138.81	-700.98
	$\beta_{1,1}$	-0.09	-0.05*	0.55	12.04	14.91	60.92*	0.55	11.99	60.95
	$\beta_{12,1}$	0.02	-0.00*	-0.43	-8.58	-10.64	-43.17*	-0.43	-8.54	-43.21
	$\beta_{1,T}$	-0.03	0.02	0.92	18.10	22.48	91.29*	0.92	18.05	91.32
	$\beta_{11,T}$	0.18	0.22	0.78	11.46	14.21	56.94*	0.77	11.42	56.97
	$\beta_{12,T}$	-0.21	-0.28	-1.42	-23.20	-28.92	-116.15	-1.41	-23.16	-116.17
	$\beta_{22,T}$	0.21	0.26	0.94	14.00	17.38	69.63*	0.94	13.96	69.67
RLS2	$\beta_{0,1}$	4.01	6.97	58.25	1058.13	1310.40	5255.61	58.69	1042.15	5250.20
	$\beta_{1,1}$	-0.38	-0.61	-4.89	-84.69	-107.92	-433.61	-4.85	-86.00	-433.14
	$\beta_{12,1}$	0.09	0.13	0.92	15.36	19.77	79.27	0.91	15.73	79.13
	$\beta_{1,T}$	-0.32	-0.54	-4.51	-78.63	-100.34	-403.23	-4.48	-79.94	-402.77
	$\beta_{11,T}$	0.20	0.24	0.99	15.29	19.03	76.35	0.98	15.27	76.36
	$\beta_{12,T}$	-0.14	-0.14	-0.07*	0.73*	1.49	6.29*	-0.08*	1.10	6.15
	$\beta_{22,T}$	0.71	1.15	8.47	152.38	187.62	751.97	8.55	149.27	751.38
GRLS	$\beta_{0,1}$	-0.01*	-0.15*	0.33*	7.37*	1.14*	-1.44*	0.54*	-0.07*	1.57*
	$\beta_{1,1}$	-0.00*	0.05*	-0.12*	-1.25*	0.03*	-0.28*	-0.13*	0.04*	-0.44*
	$\beta_{12,1}$	0.00*	-0.01	0.04*	0.03*	-0.06*	0.23*	0.04*	-0.02*	0.13*
	$\beta_{1,T}$	-0.01*	0.03*	-0.08*	-0.09*	0.07*	-0.38*	-0.07*	0.03*	-0.24*
	$\beta_{11,T}$	-0.02*	-0.07*	0.16*	3.27*	-0.08*	-0.10*	0.22*	-0.06*	0.68*
	$\beta_{12,T}$	0.03*	0.06*	-0.15	-4.14	-0.07*	0.37*	-0.24	0.05*	-0.73*
	$\beta_{22,T}$	-0.04*	-0.08*	0.19*	6.13*	0.43*	-0.88*	0.34*	-0.05*	1.01*
<u>RMSE</u>										
RLS1	$\beta_{0,1}$	0.11*	0.56*	7.41*	138.82*	174.04*	700.98*	7.40*	138.81	700.98
	$\beta_{1,1}$	0.13*	0.11*	0.55*	12.64*	15.41*	61.04**	0.64*	11.99	60.95
	$\beta_{12,1}$	0.07*	0.07*	0.43*	9.01*	10.98*	43.26*	0.47*	8.54	43.21
	$\beta_{1,T}$	0.10*	0.10*	0.92*	18.50*	22.82*	91.37*	0.97*	18.05	91.32
	$\beta_{11,T}$	0.20*	0.23*	0.78*	11.79*	14.47*	57.00**	0.80*	11.42	56.97
	$\beta_{12,T}$	0.22*	0.29	1.42	23.37	29.04	116.18*	1.43	23.16	116.19
	$\beta_{22,T}$	0.22*	0.27*	0.94*	14.27*	17.59*	69.68**	0.97*	13.96	69.67
RLS2	$\beta_{0,1}$	15.60	16.62	60.15	1239.41	1450.77	5292.23	64.15	1042.24	5250.26
	$\beta_{1,1}$	1.07	1.12	4.98	93.25	115.16	435.61	5.05	86.00	433.14
	$\beta_{12,1}$	0.28	0.28	0.95	18.40	22.35	80.01	0.98	15.72	79.13
	$\beta_{1,T}$	1.05	1.08	4.61	87.78	108.09	405.39	4.70	79.94	402.77
	$\beta_{11,T}$	0.21	0.25	0.99	15.60	19.28	76.41	1.01	15.27	76.36
	$\beta_{12,T}$	0.29	0.28*	0.26*	10.16*	10.54*	12.53*	0.39*	1.13	6.16*
	$\beta_{22,T}$	2.34	2.52	8.75	179.80	208.85	757.48	9.33	149.29	751.39
GRLS	$\beta_{0,1}$	7.69	7.76	7.72	335.58	323.92	334.33*	22.99	10.07*	23.87*
	$\beta_{1,1}$	2.36	2.31	2.33	100.12	97.47	102.39	6.58	3.04*	6.80*
	$\beta_{12,1}$	0.80	0.77	0.78	33.16	32.73	34.48*	2.03	1.01*	2.08*
	$\beta_{1,T}$	1.51	1.44	1.47	62.00	61.27	64.60*	3.72	1.87*	3.82*
	$\beta_{11,T}$	3.30	3.31	3.31	143.87	138.82	144.59	9.71	4.43*	10.03*
	$\beta_{12,T}$	3.50	3.53	3.52	153.36	147.68	152.90	10.37	4.67	10.74
	$\beta_{22,T}$	4.88	4.95	4.91	214.27	206.45	212.56	14.44	6.44*	14.99*
<u>Mean *R*-squared</u>										
RLS1		0.80	0.84	0.88	0.68	0.72	0.81	0.87	0.82	0.82
RLS2		0.81	0.85	0.90	0.71	0.76	0.84	0.89	0.86	0.85
GRLS		0.82*	0.86*	0.98*	0.83*	0.89*	0.99*	0.97*	1.00*†	1.00*†

† Strictly less than 1 when rounded to 4 decimal places.

conventional RLS estimators, and more of the variation in $\ln(y_t)$ than can be regarded as systematic, an interesting finite-sample property which is also shared by the OLS estimator in the case of the simple unrestricted linear regression model.

Columns *B* and *C* in Table 1 show the effects of increasing ρ (and, consequently, the mean of γ_t) while holding all other control parameters constant. Increasing ρ from 0.0 to 0.8 gives rise to slight departures from the sufficient conditions given by (16): for example, $E\{\beta_{1,1}\} = E\{\beta_{2,1}\}$ and $E\{\beta_{1,T}\} = E\{\beta_{2,T}\}$ depart noticeably from 0.5, the value at which the constraints given by (16) will hold exactly. The relative performance of the three estimators appears to be largely unaffected by these changes in the value of ρ: GRLS still appears to be the least biased and to have the highest explanatory power, and RLS1 still appears to have lowest RMSE. The GRLS *R*-squared tracks the true *R*-squared remarkably well, and is noticeably higher than the explanatory power of the conventional RLS estimators when $\rho = 0.8$.

In Columns *D* to *F* we examine the effects of quite radical departures from the sufficient conditions (16), by increasing the value of δ from 0.65 to 50. In Column *D* we set $\rho = \sigma_u^2 = 0$ to ensure γ_t is observation-invariant, and we set $\sigma_e^2 = 10.5$ to ensure, once again, that approximately 80% of the variation in $\ln(y_t)$ is systematic. Columns *E* and *F*, where we set $\rho = 0.2$ and $\rho = 0.8$, are used to examine the effects of further increases in the mean of γ_t. Not surprisingly, Columns *D* to *F* reveal that the conventional RLS estimators are sensitive to departures from the sufficient conditions (16): both conventional RLS estimators are noticeably biased, although the RMSE of RLS1 is still relatively small, implying RLS1 can be used to obtain estimates of β_t which will vary little in repeated samples but will be a long way from the truth. In contrast, the GRLS estimator is relatively unbiased with a relatively high RMSE, implying GRLS can be used to obtain estimates of β_t which may vary somewhat from sample to sample but will, on average, be close to the truth. Finally, the GRLS *R*-

squared still tracks the true *R*-squared extremely well, and is much higher than the conventional RLS *R*-squared values for all values of ρ. Thus, the GRLS estimator continues to dominate the conventional RLS estimators when it comes to within-sample predictive performance.

In Columns *G* to *I* we examine the effects of non-zero σ_u^2 (ie. stochastic γ_t). In Column *G* we set $\rho = 0.8$ to examine the effects of magnifying the variance of an autocorrelated γ_t; $\sigma_u^2 = \sigma_e^2 = 0.08$ to ensure that approximately 80% of the variation in $\ln(y_t)$ is systematic; and $\delta = 0.65$ to ensure that the sufficient conditions given by (16) are close to being met. In Columns *H* and *I* we set $\delta = 50$ to examine the effects of particularly large departures from the sufficient conditions (16). Column *G* reveals that, when the sufficient conditions (16) are close to being met, the relative and absolute performance of the three estimators is not significantly affected by a stochastic γ_t. However, columns *H* and *I* reveal that the relative performance of the GRLS estimator improves considerably with radical departures from the sufficient conditions, to the extent that the GRLS estimator dominates both conventional RLS estimators in terms of all three performance criteria (bias, RMSE and *R*-squared). It is interesting that the performance of the GRLS estimator of the random coefficient vector γ_t should be so good despite the fact that γ_t is not estimated within a random coefficients framework (which could be done).

Our final set of results is presented in Table 2 where we examine the effects of an increase in sample size from $T = 50$ to $T = 400$. This increase in sample size appears to accentuate the differences in estimator performance observed in Table 1. Specifically, the GRLS estimator is the least biased and, unlike the conventional RLS estimators, the degree of bias is largely unaffected by a stochastic γ_t, or by departures from the sufficient conditions given by (16). The RLS1 estimator tends to have lowest RMSE in cases where the conditions (16) are nearly

Table 2. – Monte Carlo Results for $T = 400$.

		A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
True Values										
ρ		0.00	0.20	0.80	0.00	0.20	0.80	0.80	0.00	0.80
σ_e		0.25	0.25	0.25	10.50	10.50	10.50	0.08	0.08	0.08
σ_u		0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.08	0.08
δ		0.65	0.65	0.65	50.00	50.00	50.00	0.65	50.00	50.00
CV		0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.002	0.0005
$E\{\beta_{0,1}\} = E\{\beta_{11,1}\} = E\{\beta_{0,T}\}$		0.65	0.81	3.25	50.00	62.50	250.00	3.25	50.00	250.00
$E\{\beta_{1,1}\} = E\{\beta_{2,1}\}$		0.48	0.59	2.25	34.21	42.75	170.90	2.25	34.21	170.90
$E\{\beta_{12,1}\} = E\{\beta_{22,1}\}$		0.01	-0.02	-0.47	-9.02	-11.31	-45.60	-0.47	-9.02	-45.60
$E\{\beta_{1,T}\} = E\{\beta_{2,T}\}$		0.56	0.70	2.78	42.70	53.38	213.50	2.78	42.70	213.50
$E\{\beta_{11,T}\}$		0.12	0.14	0.45	6.29	7.85	31.29	0.45	6.29	31.29
$E\{\beta_{12,T}\}$		-0.21	-0.28	-1.32	-21.20	-26.52	-106.27	-1.32	-21.20	-106.27
$E\{\beta_{22,T}\}$		0.32	0.39	1.49	22.51	28.13	112.44	1.49	22.51	112.44
R-squared		0.84	0.87	0.98	0.85	0.90	0.99	0.91	1.00†	1.00†
Bias										
RLS1	β0,1	-0.12	-0.59	-7.75	-145.12	-181.79	-732.64	-7.75	-145.09	-732.64
	β1,1	-0.09	-0.06	0.41	9.49	11.92	48.22	0.41	9.47	48.24
	β12,1	0.01*	-0.03	-0.53	-10.29	-12.86	-52.04	-0.54	-10.29	-52.01
	β1,T	-0.01*	0.05	0.94	17.99	22.55	90.82	0.94	17.98	90.84
	β11,T	0.12	0.14	0.51	7.57	9.40	37.73	0.51	7.56	37.70
	β12,T	-0.21	-0.28	-1.38	-22.48	-28.07	-112.71	-1.38	-22.47	-112.68
	β22,T	0.31	0.39	1.55	23.78	29.68	118.87	1.55	23.77	118.84
RLS2	β0,1	3.99	7.50	63.41	1140.00	1428.66	5725.90	63.37	1134.52	5719.21
	β1,1	-0.25	-0.39	-2.62	-45.96	-57.24	-229.11	-2.63	-45.52	-228.97
	β12,1	0.05	0.06	0.27	4.34	5.41	21.19	0.27	4.24	21.18
	β1,T	-0.17	-0.28	-2.10	-37.46	-46.61	-186.51	-2.10	-37.02	-186.37
	β11,T	0.12	0.15	0.54	8.09	10.06	40.36	0.54	8.08	40.33
	β12,T	-0.16	-0.19	-0.58	-7.84	-9.81	-39.48	-0.57	-7.95	-39.48
	β22,T	0.81	1.37	10.00	175.90	220.55	884.39	9.90	175.40	883.43
GRLS	β0,1	0.08*	0.00*	-0.02*	2.49*	0.13*	-1.24*	-0.19*	-0.01*	-0.05*
	β1,1	-0.03*	0.00*	0.02*	-0.92*	0.10*	0.43*	0.07*	-0.01*	0.03*
	β12,1	0.01*	-0.00*	-0.02*	0.39*	0.06*	-0.15*	-0.03*	0.01*	-0.02*
	β1,T	-0.02	0.00*	0.02*	-0.65*	-0.04*	0.15*	0.04*	-0.00*	0.03*
	β11,T	0.02*	0.00*	-0.00*	0.58*	-0.21*	-0.40*	-0.06*	-0.01*	-0.02*
	β12,T	-0.03*	-0.01*	-0.01*	-0.65*	0.38*	0.65*	0.09*	0.01*	0.02*
	β22,T	0.05*	0.01*	0.01*	1.07*	-0.62*	-1.17*	-0.15*	-0.03*	-0.02*
RMSE										
RLS1	β0,1	0.12	0.59*	7.75	145.12	181.79	732.64	7.75*	145.09	732.64
	β1,1	0.09	0.06*	0.41*	9.52*	11.94*	48.22	0.42*	9.47	48.24
	β12,1	0.02	0.03*	0.53	10.31	12.87	52.04	0.54	10.29	52.01
	β1,T	0.02	0.05*	0.94	18.00	22.56	90.82	0.96*	17.98	90.84
	β11,T	0.12*	0.15*	0.51*	7.59*	9.43*	37.73	0.52*	7.56	37.70
	β12,T	0.21	0.28	1.38	22.48	28.08	112.71	1.38	22.47	112.68
	β22,T	0.31	0.39*	1.55	23.79*	29.69*	118.88	1.56*	23.77	118.85
RLS2	β0,1	5.94	8.73	63.56	1155.64	1440.71	5729.13	64.71	1134.53	5719.23
	β1,1	0.35	0.46	2.64	47.08	58.18	229.33	2.68	45.53	228.97
	β12,1	0.08	0.09	0.27	5.12*	6.05*	21.36	0.29*	4.24	21.18
	β1,T	0.30	0.37	2.11	38.83	47.77	186.78	2.16	37.02	186.37
	β11,T	0.12*	0.15*	0.54	8.12	10.08	40.36	0.55	8.08	40.33
	β12,T	0.18*	0.20*	0.58*	8.29*	10.18*	39.57	0.59*	7.95	39.48
	β22,T	1.05	1.53	10.02	178.21	222.32	884.86	10.20	175.41	883.43
GRLS	β0,1	1.94	1.91	1.90*	79.86*	80.71*	79.35*	8.20	3.04*	8.22*
	β1,1	0.73	0.73	0.72	29.92	30.65	29.85*	2.74	1.12*	2.76*
	β12,1	0.26	0.26	0.26*	10.69	11.10	10.62*	0.89	0.41*	0.90*
	β1,T	0.43	0.43	0.42*	17.34*	18.05*	17.22*	1.20	0.62*	1.22*
	β11,T	0.59	0.59	0.58	24.45	24.58	24.37*	2.48	0.94*	2.49*
	β12,T	0.88	0.88	0.87	36.86	36.62	36.65*	3.89	1.41*	3.90*
	β22,T	1.56	1.55	1.54*	65.44	64.74	64.95*	7.01	2.49*	7.02*
Mean *R*-squared										
RLS1		0.84*	0.87*	0.91	0.74	0.78	0.85	0.84	0.86	0.86
RLS2		0.84*	0.87*	0.92	0.76	0.80	0.88	0.86	0.89	0.89
GRLS		0.84*	0.87*	0.98*	0.86*	0.90*	0.99*	0.91*	1.00*†	1.00*†

† Strictly less than 1 when rounded to 4 decimal places.

satisfied, but GRLS begins to dominate in terms of RMSE when departures from these conditions become large, and when γ_t is stochastic. Finally, the GRLS estimator clearly dominates the conventional RLS estimators in terms of within-sample predictive performance.

V. Conclusions

The estimator used most frequently to impose observation-varying equality constraints on the parameters of linear models is unsatisfactory – because the models' parameters are typically assumed to be observation-invariant, Restricted Least Squares (RLS) unnecessarily underconstrains or overconstrains the parameter space. We show how to overcome the problem by relaxing the assumption that the parameters are fixed.

It is possible to impose observation-varying equality constraints on the observation-varying parameters of linear models in several ways. For example, Doran and Rambaldi (1997) use the constraints to augment the observation equation in a state-space model, and then estimate the parameters using the Kalman filter. Unfortunately, the practical usefulness of this approach is limited by the fact that the procedure involves the estimation of a full covariance matrix, and optimisation routines tend to encounter convergence problems when the number of parameters/elements is large. In contrast, O'Donnell, Shumway and Ball (1999) use a more conventional approach involving direct substitution of the constraints into the econometric model. Unfortunately, the problem with their substitution method is that the parameter estimates are not invariant to an arbitrary re-ordering of the regressors. In this paper we overcome this problem – we show how to substitute the constraints into the model in such a way that the parameter estimates do not depend on the ordering of the regressors. The resulting estimator has several additional desirable properties, including the fact that it collapses to conventional RLS in the special case where the equality constraints are observation-invariant. Our so-called Generalised Restricted Least Squares (GRLS) estimator is a computationally simple estimator that allows us to impose observation-varying equality constraints in the *exact* form economic theory prescribes.

The computational simplicity of our estimator has enabled us to conduct a reasonably extensive Monte Carlo investigation of its properties in small and large samples. Overall, these simulation results suggest that GRLS is robust to different types of data generating processes, and superior to conventional RLS estimators in terms of those performance criteria which can be regarded as being most important for empirical work, namely bias and within-sample predictive performance. Although GRLS appears inferior to RLS in terms of RMSE, we discount the practical importance of the RMSE criterion on the grounds that it is better to be vaguely right (small bias with large variance) than precisely wrong (large bias with small variance). We also discount the practical usefulness of one of the most commonly used RLS estimators (the estimator which imposes sufficient but not necessary conditions for the restrictions implied by theory to hold) on the grounds that sufficient conditions do not always exist. Even when they do exist, there is no theoretical reason for believing that they appropriately restrict the parameters. Our experiments show that when such restrictions are inappropriate, large biases occur. Of course, the other RLS estimator (the estimator which imposes restrictions at mean values of the data) is of little practical value because the constraints are only imposed at one point, and the experiments conducted in this paper indicate that its performance is uniformly poor.

REFERENCES

Beattie, B.C., Taylor, C.R., 1985. *The Economics of Production*. Wiley, New York.

Clements, K.W., Izan, H.Y., 1987. The measurement of inflation: a stochastic approach. *Journal of Business and Economic Statistics* 5, 339-359.

Doran, H.E., Rambaldi, A.N., 1997. Applying linear time-varying constraints to econometric models: With an application to demand systems. *Journal of Econometrics* 79, 83-95.

Graybill, F.A. 1969. *Introduction to Matrices with Applications in Statistics*. Wadsworth, Belmont CA.

Goldberger, A.S.1964. *Econometric Theory*. Wiley, New Your

Greene, W.H. 2000. *Econometric Analysis*. 4th ed. Prentice Hall, New Jersey.

Judge, G.G,W.E. Griffiths, R.C. Hill, H. Lütkepohl, and T-C Lee. 1985. *The theory and Practice of Econometrics*. 2nd ed. Wiley, New York.

Kalaba, R., Tesfatsion, L. 1989. Time-Varying Linear Regression via Flexible Least Squares. *Computers and Mathematics with Applications*, 17:1215-1245.

O'Donnell,C.J., Shumway, C.R., Ball, V.E. 1999. Input Input Demands and Inefficiency in U.S. Agriculture. *American Journal of Agricultural Economics*, 81(November), 865-880.

Selvanathan, E.A., 1989. A note on the stochastic approach to index numbers. *Journal of Business and Economic Statistics* 7, 471-474.

APPENDIX

Proofs

P.1: Let $\boldsymbol{R}_t = \boldsymbol{R}$ and $\mathbf{r}_t = \mathbf{r}$ for all t. Then the estimator defined by (10) and (11) is identical to the RLS estimator $\mathbf{b}_{\mathrm{RLS}} = \mathbf{b} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\mathbf{b} - \mathbf{r})$ where $\boldsymbol{X} = (\boldsymbol{X}_1', ..., \boldsymbol{X}_T')'$, $\mathbf{y} = (\mathbf{y}_1', ..., \mathbf{y}_T')'$ and $\mathbf{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathbf{y}$.

Proof: Define $\mathbf{y} = (\mathbf{y}_1', ..., \mathbf{y}_T')'$ and $\boldsymbol{X} = (\boldsymbol{X}_1', ..., \boldsymbol{X}_T')'$. If $\boldsymbol{R}_t = \boldsymbol{R}$ and $\mathbf{r}_t = \mathbf{r}$ for all t then

(A.1) $$\boldsymbol{H} = \boldsymbol{I}_K - \boldsymbol{R}^+\boldsymbol{R}.$$

(A.2) $$\boldsymbol{Z} = \boldsymbol{X}(\boldsymbol{I}_K - \boldsymbol{R}^+\boldsymbol{R}) = \boldsymbol{X}\boldsymbol{H},$$

(A.3) $$\mathbf{w} = \mathbf{y} - \boldsymbol{X}\boldsymbol{R}^+\mathbf{r},$$

and $\mathbf{g}$ solves the normal equations $(\boldsymbol{Z}'\boldsymbol{Z})\mathbf{g} = \boldsymbol{Z}'\mathbf{w}$. The RLS estimate $\mathbf{b}_{\mathrm{RLS}}$ is unique and satisfies the equations $\boldsymbol{R}\mathbf{b}_{\mathrm{RLS}} = \mathbf{r}$ and $(\boldsymbol{X}'\boldsymbol{X})\mathbf{b}_{\mathrm{RLS}} = \boldsymbol{X}'\mathbf{y} + \boldsymbol{R}'\lambda$ where λ is an undetermined vector. By construction, the GRLS estimate $\mathbf{b} = \boldsymbol{R}^+\mathbf{r} + \boldsymbol{H}\mathbf{g}$ satisfies $\boldsymbol{R}\mathbf{b} = \mathbf{r}$. It follows that $\mathbf{b} = \mathbf{b}_{\mathrm{RLS}}$ if we can find a vector λ such that $(\boldsymbol{X}'\boldsymbol{X})\mathbf{b} = \boldsymbol{X}'\mathbf{y} + \boldsymbol{R}'\lambda$. Consider

(A.4) $$\begin{aligned}\boldsymbol{Z}'\boldsymbol{X}\mathbf{b} &= \boldsymbol{Z}'\boldsymbol{X}\boldsymbol{R}^+\mathbf{r} + \boldsymbol{Z}'\boldsymbol{X}\boldsymbol{H}\mathbf{g} \\ &= \boldsymbol{Z}'\boldsymbol{X}\boldsymbol{R}^+\mathbf{r} + (\boldsymbol{Z}'\boldsymbol{Z})\mathbf{g} \\ &= \boldsymbol{Z}'\boldsymbol{X}\boldsymbol{R}^+\mathbf{r} + \boldsymbol{Z}'\mathbf{w} \\ &= \boldsymbol{Z}'[\boldsymbol{X}\boldsymbol{R}^+\mathbf{r} + (\mathbf{y} - \boldsymbol{X}\boldsymbol{R}^+\mathbf{r})] \\ &= \boldsymbol{Z}'\mathbf{y}.\end{aligned}$$

Therefore $\boldsymbol{Z}'(\boldsymbol{X}\mathbf{b} - \mathbf{y}) = (\boldsymbol{I}_K - \boldsymbol{R}^+\boldsymbol{R})[(\boldsymbol{X}'\boldsymbol{X})\mathbf{b} - \boldsymbol{X}'\mathbf{y}] = \mathbf{0}$. Thus, $(\boldsymbol{X}'\boldsymbol{X})\mathbf{b} - \boldsymbol{X}'\mathbf{y}$ lies in the orthogonal complement of $(\boldsymbol{I}_K - \boldsymbol{R}^+\boldsymbol{R})$. That is, it lies in the space spanned by the columns of $\boldsymbol{R}'$ (see Graybill, Theorem 6.4.11, p.107). Therefore, for some λ, $(\boldsymbol{X}'\boldsymbol{X})\mathbf{b} - \boldsymbol{X}'\mathbf{y} = \boldsymbol{R}'\lambda$ and the result $\mathbf{b} = \mathbf{b}_{\mathrm{RLS}}$ is proved.

P.2: Let $\boldsymbol{R}_t = [\boldsymbol{R}_{1t}, \mathbf{0}_{J,K_2}]$ where $\boldsymbol{R}_{1t}$ is $J \times K_1$ ($J < K_1$) of rank J and $K_2 \equiv K - K_1$. Let β_t and $\mathbf{b}_t$ be partitioned conformably as $\beta_t = [\beta_{1t}', \beta_{2t}']'$ and $\mathbf{b}_t = [\mathbf{b}_{1t}', \mathbf{b}_{2t}']'$. Then $\mathbf{b}_{2t}$ is always observation-invariant.

Proof: If $\boldsymbol{R}_t = [\boldsymbol{R}_{1t}, \mathbf{0}_{J,K_2}]$ then it is easily verified that

(A.5) $$\boldsymbol{R}_t^+ = \begin{bmatrix} \boldsymbol{R}_{1t}^+ \\ \mathbf{0}_{K_2,J} \end{bmatrix}$$

and

(A.6) $$\mathbf{b}_t = \begin{bmatrix} \boldsymbol{R}_{1t}^+ \\ \mathbf{0}_{K_2,J} \end{bmatrix}\mathbf{r}_t + (\boldsymbol{I}_K - \begin{bmatrix} \boldsymbol{R}_{1t}^+ \\ \mathbf{0}_{K_2,J} \end{bmatrix}[\boldsymbol{R}_{1t}^+, \mathbf{0}_{J,K_2}])\mathbf{g}$$

$$= \begin{bmatrix} \boldsymbol{R}_{1t}^+\mathbf{r}_t \\ \mathbf{0}_{K_2,1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{I}_{K_1} - \boldsymbol{R}_{1t}^+\boldsymbol{R}_{1t} & \mathbf{0}_{K_1,K_2} \\ \mathbf{0}_{K_2,K_1} & \boldsymbol{I}_{K_2} \end{bmatrix}\begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix}$$

where $\mathbf{g}$ has been partitioned conformably as $\mathbf{g} = [\mathbf{g}_1', \mathbf{g}_2']'$. It follows that $\mathbf{b}_{2t} = \mathbf{g}_2$ and is observation-invariant.

P.3: The estimator defined by (10) and (11) is invariant with respect to a re-ordering of the regressors.

Proof: A re-ordering of the regressors in (1) yields

(A.7) $$\mathbf{y}_t = \boldsymbol{X}_t^* \beta_t^* + \mathbf{e}_t$$

where $\boldsymbol{X}_t^* = \boldsymbol{X}_t \boldsymbol{W}$, $\beta_t^* = \boldsymbol{W}'\beta_t$ and $\boldsymbol{W}$ is a permutation (orthogonal) matrix. A necessary and sufficient condition for the elements of $\mathbf{b}_t$ to be invariant under a re-ordering of the regressors is that the GRLS estimate of β_t^* is $\mathbf{b}_t^* = \boldsymbol{W}'\mathbf{b}_t$. To see this holds, note that the constraint equation corresponding to the re-ordered model (A.7) is

(A.8) $$\boldsymbol{R}_t^* \beta_t^* = \mathbf{r}_t$$

where $\boldsymbol{R}_t^* = \boldsymbol{R}_t \boldsymbol{W}$ and $\boldsymbol{R}_t^{*+} = (\boldsymbol{R}_t \boldsymbol{W})^+ = \boldsymbol{W}'\boldsymbol{R}_t^+$ (Graybill, Theorem 6.2.10, p.100). The general solution to (A.8) is

(A.9) $$\beta_t^* = \boldsymbol{R}_t^{*+}\mathbf{r}_t + (\boldsymbol{I}_K - \boldsymbol{R}_t^{*+}\boldsymbol{R}_t^*)\gamma_t^*$$

where γ_t^* is an arbitrary $K \times 1$ vector. Substituting (A.9) into (A.7) yields

(A.10) $$\mathbf{w}_t^* = \boldsymbol{Z}_t^* \gamma_t^* + \mathbf{e}_t$$

where

(A.11) $$\mathbf{w}_t^* = \mathbf{y}_t - \boldsymbol{X}_t^* \boldsymbol{R}_t^{*+} \mathbf{r}_t$$

$$= \mathbf{y}_t - \boldsymbol{X}_t\boldsymbol{W}\boldsymbol{W}'\boldsymbol{R}_t^{+}\mathbf{r}_t$$

$$= \mathbf{w}_t$$

and

$$\text{(A.12)} \qquad \boldsymbol{Z}_t^* = \boldsymbol{X}_t^*(\boldsymbol{I}_K - \boldsymbol{R}_t^{*+}\boldsymbol{R}_t^*)$$

$$= \boldsymbol{X}_t\boldsymbol{W}(\boldsymbol{I}_K - \boldsymbol{W}'\boldsymbol{R}_t^{+}\boldsymbol{R}_t\boldsymbol{W})$$

$$= \boldsymbol{X}_t\boldsymbol{W}\boldsymbol{W}'(\boldsymbol{I}_K - \boldsymbol{R}_t^{+}\boldsymbol{R}_t)\boldsymbol{W}$$

$$= \boldsymbol{X}_t(\boldsymbol{I}_K - \boldsymbol{R}_t^{+}\boldsymbol{R}_t)\boldsymbol{W}$$

$$= \boldsymbol{Z}_t\boldsymbol{W}.$$

Under the invariance assumption $\gamma_t^* = \gamma^*$ we obtain the analogue of (11) as

$$\text{(A.13)} \qquad \mathbf{g}^* = (\boldsymbol{Z}^{*}{}'\boldsymbol{Z}^*)^{-1}\boldsymbol{Z}^{*}{}'\mathbf{w}^*$$

$$= (\boldsymbol{W}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{Z}'\mathbf{w}$$

$$= \boldsymbol{W}'\mathbf{g}$$

where $\mathbf{w}^* = (\mathbf{w}_1^{*}{}', \dots, \mathbf{w}_T^{*}{}')'$ and $\boldsymbol{Z}^* = (\boldsymbol{Z}_1^{*}{}', \dots, \boldsymbol{Z}_T^{*}{}')'$. Finally, the analogue of (10) is

$$\text{(A.9)} \qquad \mathbf{b}_t^* = \boldsymbol{R}_t^{*+}\mathbf{r}_t + (\boldsymbol{I}_K - \boldsymbol{R}_t^{*+}\boldsymbol{R}_t^*)\mathbf{g}^*$$

$$= \boldsymbol{W}'\boldsymbol{R}_t^{+}\mathbf{r}_t + (\boldsymbol{I}_K - \boldsymbol{W}'\boldsymbol{R}_t^{+}\boldsymbol{R}_t\boldsymbol{W})\boldsymbol{W}'\mathbf{g}$$

$$= \boldsymbol{W}'[\boldsymbol{R}_t^{+}\mathbf{r}_t + (\boldsymbol{I}_K - \boldsymbol{R}_t^{+}\boldsymbol{R}_t)\boldsymbol{W}\boldsymbol{W}'\mathbf{g}]$$

$$= \boldsymbol{W}'[\boldsymbol{R}_t^{+}\mathbf{r}_t + (\boldsymbol{I}_K - \boldsymbol{R}_t^{+}\boldsymbol{R}_t)\mathbf{g}]$$

$$= \boldsymbol{W}'\mathbf{b}_t$$

as required.