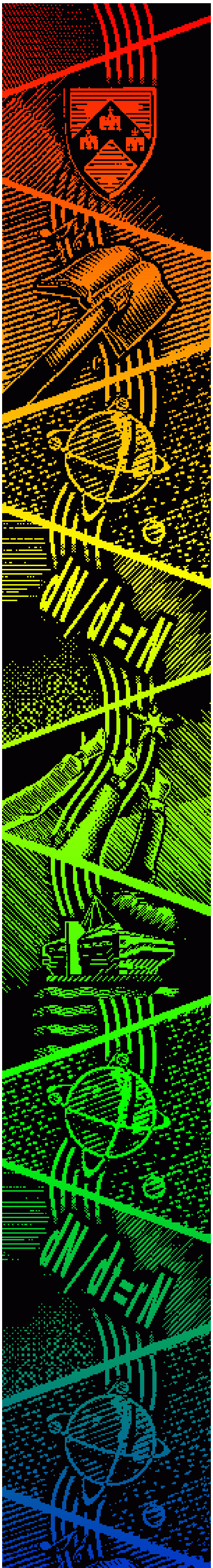

THE UNIVERSITY *of York*

Discussion Papers in Economics

No. 2005/11

Structural Determinants of Cumulative Endogeneity Bias

by

David Mayston

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

STRUCTURAL DETERMINANTS

OF

CUMULATIVE ENDOGENEITY BIAS

by

DAVID MAYSTON*

University of York

* The author is grateful to two anonymous referees of the *Journal of Multivariate Analysis* for helpful comments on an earlier draft of this paper.

ABSTRACT

The BLU properties of OLS estimators under known assumptions have encouraged the widespread use of OLS multivariate regression analysis in many empirical studies that are based upon a conceptual model of a single explanatory equation. However, such a model may well be an imperfect empirical approximation to the valid underlying conceptual model, that may well contain several important additional interrelationships between the relevant variables. In this paper, we examine the conditions under which we can predict the direction of the resultant endogeneity bias that will prevail in the OLS asymptotic parameter estimates for any given endogenous or predetermined variable, and the extent to which we can rely upon simple heuristics in this process. We also identify the underlying structural parameters to which the magnitude of the endogeneity bias is sensitive. The importance of such sensitivity analysis has been underlined by an increasing awareness of the inability of standard diagnostic tests to shed light upon the *extent* of the endogeneity bias, rather than upon merely its *existence*. The paper examines the implications of the analysis for statistical inferences about the true value of the regression coefficients and the validity of associated t-statistics.

Keywords: Multivariate analysis and parameter estimation; Cumulative endogeneity bias.

(June 2005, Revised October 2007)

1. Introduction

The best linear unbiased (BLU) properties of ordinary least-squares (OLS) under known assumptions have encouraged the widespread use of OLS multivariate regression analysis in many empirical studies that are based upon a conceptual model of a single explanatory equation. However, such a model may well be an imperfect empirical approximation to the valid underlying conceptual model, which may contain several important additional interrelationships between the relevant variables. These additional interrelationships may well undermine the strict exogeneity assumption [2] used to generate the standard BLU properties of OLS, so that the OLS parameter estimates may be biased even as the sample size increases asymptotically, and hence be no longer consistent. However, estimation techniques, such as Instrumental Variables (IV), that are intended to overcome OLS endogeneity bias, involve requirements, such as the availability of a whole set of suitable instruments that are uncorrelated with the disturbance term but strongly correlated with the original variable, which may not hold in practice. As we demonstrate later in the paper, rather than overcoming it, the substitution of instruments as proxy variables for only a subset of the relevant endogenous variables may instead *increase* the magnitude of the endogeneity bias, even when these instruments are uncorrelated with the disturbance term in the primary equation of interest. It is therefore important for both statistical theorists and empiricists to understand more fully the factors which influence both the direction and the magnitude of the endogeneity bias which may result from OLS multivariate estimation.

As Nakamura and Nakamura [11, 12] have stressed, standard empirical tests [6, 14] for the *existence* of such OLS bias do not perform well as predictors of the *extent* of such bias, despite the fact that it is the extent of the bias which may be of prime interest in many applied studies. Moreover, as Magnus and Vasnev [9] have recently emphasised, standard diagnostic tests may similarly provide little or no information about the *sensitivity* of parameter estimates to departures from the underlying assumptions of the model that is being estimated, but instead may indeed be (asymptotically) statistically independent from such sensitivity assessments. Diagnostic testing alone, upon which econometric practice has tended to concentrate, will therefore do little to fill the vacuum left by the relative neglect of the important questions which fall within the domain of *sensitivity analysis*. These questions include the nature of the sensitivity of the extent and direction of the OLS endogeneity bias to key features of the conceptual model that adequately characterises the underlying interrelationship between the variables.

Being able to assess the direction of bias associated with existing OLS studies, and the factors to which its magnitude may be sensitive, is of substantial potential value in many policy and decision-making contexts. There are, for example, many existing studies (see [4, 5]) based upon OLS estimation of the effectiveness of resources in boosting educational performance, which yield parameter estimates of this effectiveness that appear to be not significantly different from zero. If these estimates were taken at their face value, they would have important implications for policy decisions, such as that allocating additional resources to the educational sector cannot be justified by their being expected to have a positive impact on educational performance. In interpreting the results of such existing OLS multivariate studies, there is a need to assess the likely direction of the bias that these estimates may involve, and the conditions under which this bias may be large.

It is therefore of considerable interest in many decision contexts to be able to understand the conditions under which:

a. multiple additional relationships will pull the cumulative asymptotic bias in the estimate of any given coefficient of the primary equation in an overall direction that can be predicted from insights into the qualitative characteristics of the underlying structural parameters of the model that these multiple additional relationships and the primary equation generate;

b. simple heuristic rules can assist in determining the overall direction of the cumulative endogeneity bias;

c. any given additional relationship adds to, rather than offsets, the overall strength of the cumulative asymptotic bias;

d. the sign and magnitude of some structural parameters of the model are irrelevant to determining the cumulative asymptotic bias for a given coefficient;

e. an (upper or lower) bound upon the true value of the regression coefficient can be derived from its OLS asymptotic estimate;

f. the true value of the regression coefficient is an integer multiple of its OLS asymptotic estimate;

g. an unbiased estimate of the standard t-statistic for assessing the significance of a given coefficient estimate would attain the level conventionally associated with a given degree of statistical significance, even though a biased OLS estimate of the t-statistic appears not to be statistically significant;

h. the impact of the omission of some endogenous and/or predetermined variables from the OLS regression upon the direction and extent of the cumulative asymptotic bias can be predicted.

The paper is organised as follows. Section 2 focuses on the conditions under which it is possible to determine the direction of the cumulative endogeneity bias. Section 3 examines the implications of the analysis for drawing statistical inferences about the true value of the regression coefficients in the presence of endogeneity bias. Section 4 considers the sensitivity of the cumulative endogeneity bias to changes in the value of key parameters. Section 5 extends the analysis to cases where some predetermined and/or endogenous variables are not included as regressors in the OLS estimation, and examines the impact of replacing an endogenous variable by an instrumental variable upon the extent of the cumulative bias. Section 6 concludes our discussion, with the mathematical proofs contained in the Appendix.

2. The Generation of Cumulative Endogeneity Bias

The primary equation of interest will be assumed to be of the form:

$$x_{i1} = \sum_{k=2}^{n} \beta_k x_{ik} + \sum_{h=1}^{m} \beta_{n+h} z_{ih} - u_{i1} \tag{1}$$

However, there also exist multiple additional inter-relationships of the form:

$$x_{ij} = \sum_{\substack{k=1 \\ k\neq j}}^{n} b_{kj} x_{ik} + \sum_{h=1}^{m} c_{hj} z_{ih} - u_{ij} \quad for \;\; j = 2,....,n \tag{2}$$

where across a sample of p observations, denoted by i = 1, ..., p, x_{ik} denotes the ith observation on the kth variable that is endogenous to the inter-relationships (1) - (2) for k = 1,..., n, and z_{ih} denotes the ith observation on the hth predetermined variable that is not endogenously determined by the inter-relationships (1) - (2). Both the x_{ik} and the z_{ih} will be assumed to be expressed in terms of deviations from their respective sample means. The b_{kj} and c_{hj} are the corresponding structural parameters in the jth inter-relationship, with

$$b_{k1} \equiv \beta_k \ \ for\ k \in J_1 \equiv \{2,...,n\} \ \ and \ \ c_{h1} \equiv \beta_{n+h} \ \ for\ h \in M \equiv \{1,...,m\} \tag{3}$$

u_{ij} is the random disturbance term for the ith observation in the jth structural relation in the model (1) – (2). The u_{ij} are assumed to be independently and identically multivariate normally distributed for each observation i = 1,...., p, with zero means and a covariance matrix $\boldsymbol{V} \equiv [\sigma_{kj}]$ that is symmetric and positive definite. The model implied by (1) - (3) may be written in matrix form as:

$$\mathbf{X}\boldsymbol{B} + \boldsymbol{ZC} = \boldsymbol{U} \tag{4}$$

where $\boldsymbol{X}$ is the p x n matrix with elements x_{ik}, $\boldsymbol{Z}$ is the p x m matrix with elements z_{ih}, $\boldsymbol{B}$ is an n x n matrix with elements b_{kj}, where $b_{kk} \equiv -1$ for each $k \in J \equiv \{1,...,n\}$, and $\boldsymbol{C}$ is the m x n matrix with elements c_{hj}. $\boldsymbol{U}$ is the p x n matrix with random elements u_{ij}. We will require $\boldsymbol{B}$ to be non-singular. $\boldsymbol{X_0}$ will denote the p x (n - 1) sub-matrix of $\boldsymbol{X}$ with elements x_{ik} $for\ k \in J_1$, with $\boldsymbol{Y}_0 \equiv [\boldsymbol{X_0}, \boldsymbol{Z}]$. Φ will denote the null set, with:

$$J_0 \equiv \{ j \in J_1 : b_{1j} \neq 0\}, \ \ K \equiv \{2,..,n+m\}, \ \ b_{n+h,j} \equiv c_{hj} \ \ for\ h \in M,\ j \in J \tag{5}$$

The OLS estimates of the coefficients β_k in (1) will be denoted by $\hat{\beta}_k$ for each $k \in K$, with *plim* denoting the probability limit of the entity in question, as the sample size p becomes infinitely large, and with $\hat{\beta}_k^o \equiv plim\hat{\beta}_k$. The shorthand 'iff' will denote 'if and only if'.

For this general formulation, we can establish the following Propositions, with their derivation given in the Appendix.

Proposition 1. The cumulative (asymptotic) bias, θ_k, in each estimated coefficient β_k under OLS is given by:

$$\theta_k \equiv plim\ \hat{\beta}_k - \beta_k = -[\sum_{j \in J_1}(b_{kj} + \beta_k b_{1j})\sum_{\ell \in J_1} d_{j\ell}(b_{1\ell}\sigma_{11} + \sigma_{1\ell})] \ \ for\ k = 2,...,n+m \tag{6}$$

where $[d_{j\ell}] \equiv \boldsymbol{E}^{-1}$ for $\boldsymbol{E} \equiv [e_{j\ell}]$ and $e_{j\ell} \equiv b_{1j}(b_{1\ell}\sigma_{11} + \sigma_{1\ell}) + \sigma_{j1}.b_{1\ell} + \sigma_{j\ell}$ for $j, \ell \in J_1$ (7)

Proposition 1 shows how the extent of the cumulative bias, in each endogenous variable and in each predetermined variable, depends upon the underlying model parameters, for the general case where the covariance matrix $\boldsymbol{V}$ is not necessarily diagonal. For any endogenous or predetermined variable k = 2,…, n + m, the direction and magnitude of the cumulative bias θ_k depends upon each b_{kj}, β_k and b_{1j} for all $j = 2,....,n$, and upon all the elements of the covariance matrix $\boldsymbol{V}$. However, for any given set of endogenous and predetermined variables and given values of the elements of the covariance matrix $\boldsymbol{V}$, the extent of the cumulative bias θ_k does not depend upon the values of the coefficients β_h and b_{hj} for $h(\neq k) = 2,...,n+m,$ and $j = 2,....,n$.

Even within this general context, it is possible to establish necessary and sufficient conditions on the underlying model parameters to ensure that the sign of the overall cumulative bias is determinate from (6) and (7), as in the following Proposition.

Proposition 2. The cumulative bias, θ_k, in the estimated coefficient β_k under OLS will be negative *(positive)* if there is a subset J_1' of $J_1 \equiv \{2,....,n\}$, with $J_1'' \equiv J_1 - J_1'$, such that for each $e_{j\ell}$ defined by (7) above:

$$e_{j\ell} \leq 0 \;\; for \;\; j \neq \ell \;\; whenever \;\; j,\ell \in J_1' \;\; or \;\; j,\ell \in J_1'' \tag{8}$$

$$e_{j\ell} \geq 0 \;\; for \;\; j \neq \ell \;\; whenever \;\; j \in J_1' \,\&\, \ell \in J_1'' or \;\; j \in J_1'' \& \, \ell \in J_1' \tag{9}$$

$$(b_{1\ell}\sigma_{11} + \sigma_{1\ell})(b_{kj} + \beta_k b_{1j}) \geq (\leq) \, 0 \;\; whenever \;\; j,\ell \in J_1' \;\; or \;\; j,\ell \in J_1'' \tag{10}$$

$$(b_{1\ell}\sigma_{11} + \sigma_{1\ell})(b_{kj} + \beta_k b_{1j}) \leq (\geq) \, 0 \;\; whenever \;\; j \in J_1' \,\&\, \ell \in J_1'' or \;\; j \in J_1'' \& \, \ell \in J_1' \tag{11}$$

where (10) or (11) holds as a strict inequality for at least one $\ell \in J_1$ and at least one $j \in J_1$, and E in (7) is indecomposable. Moreover, the cumulative bias becomes more negative *(positive)* for each j,ℓ combination for which (10) or (11) holds as a strict inequality. Conditions (8) – (11) are also necessary for the sign of the overall cumulative bias to be determinate from (6) and (7), given only the sign pattern of each $e_{j\ell}$, $(b_{1\ell}\sigma_{11} + \sigma_{1\ell})$ and $(b_{kj} + \beta_k b_{1j})$ for $k, j, \ell \in J_1$, whenever $n \neq 3$, $e_{j\ell} \neq 0$ for all $j,\ell \in J_1$, and $\boldsymbol{V}$ is not diagonal. In these necessary and sufficient conditions, we may have $J_1' = J_1$, with $J_1'' = \Phi$.

As we show in the Appendix, $\boldsymbol{E}$ is a positive definite matrix, with therefore $|\boldsymbol{E}| > 0$ and $e_{jj} > 0$ for all $j \in J_1$. For the case of a single endogenous variable, and hence n = 2, (8) and (9) have no force, with Proposition 2 implying:

$$\theta_k < (=,>) \, 0 \;\; \text{iff} \;\; (b_{12}\sigma_{11} + \sigma_{12})(b_{k2} + \beta_k b_{12}) > (=,<) \, 0 \;\; for \;\; k = 2,..,m+2 \tag{12}$$

In the case of two endogenous variables, with therefore n = 3, conditions (8) – (11) are sufficient for the sign of the cumulative bias θ_k to be determinate for each k = 2,...,3 + m, but not necessary. Since $|\boldsymbol{E}| > 0$, the sign pattern of $[d_{j\ell}] = \boldsymbol{E}^{-1}$ can be determined from the sign pattern of each $e_{j\ell}$ without imposing conditions (8) and (9) in the 2x2 case where n - 1 = 2. For the case of n = 3, we may derive from Proposition 1 that the more general necessary and sufficient conditions for being able to sign $\theta_k < 0$ for any variable $k = 2,...,3+m$ are that:

$$(b_{1j}\sigma_{11} + \sigma_{12})(b_{kj} + \beta_k b_{1j}) \geq 0 \ \ for \ \ j = 2,3 \tag{13}$$

$$(b_{12}\sigma_{11} + \sigma_{12})(b_{k3} + \beta_k b_{13})e_{23} \leq 0 \ \& \ (b_{13}\sigma_{11} + \sigma_{13})(b_{k2} + \beta_k b_{12})e_{32} \leq 0 \tag{14}$$

with at least one of the inequalities in (13) and (14) being a strict inequality. Similarly, for the case where n = 3, necessary and sufficient conditions for being able to sign $\theta_k > 0$ for any $k = 2,...,m+3$ are that the inequalities in (13) and (14) hold in reverse, with at least one of these inequalities being a strict inequality. For cases where there are three or more endogenous variables, and hence where n > 3, the sign pattern of $\boldsymbol{E}^{-1}$ cannot in general be determined without imposing conditions (8) and (9), unless $\boldsymbol{V}$ is diagonal. Conditions (8) – (11) then provide both necessary and sufficient conditions within Proposition 2 for being able to determine the sign of θ_k for any endogenous or predetermined variable k = 2,..., m + n when $\boldsymbol{V}$ is not diagonal, given only the sign pattern of the relevant combinations of the underlying parameters.

In the following Proposition, we examine the relationship between the standard error of the OLS estimate of the first equation under endogeneity bias and the underlying variance of the disturbance term of the first equation.

Proposition 3. The asymptotic value of $s^2 \equiv \hat{\boldsymbol{v}}'\hat{\boldsymbol{v}}/(p-n)$ (where s is the standard error of the OLS estimate of the first equation, the residuals are given by $\hat{\boldsymbol{v}} \equiv \boldsymbol{x} - \boldsymbol{Y_o}\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is the vector of OLS estimates of $\boldsymbol{\beta}$, and $\hat{\boldsymbol{v}}'\hat{\boldsymbol{v}}$ is the OLS residual sum of squares) is strictly less than the variance $\sigma_1^2 \equiv \sigma_{11}$ of the disturbance term of the first equation by the positive amount:

$$q = \sum_{j \in J_1} \sum_{\ell \in J_1} (b_{1j}\sigma_{11} + \sigma_{1j})\, d_{j\ell} (b_{1\ell}\sigma_{11} + \sigma_{1\ell}) \tag{15}$$

whenever $\sigma_{1j} \neq -b_{1j}\sigma_{11}$ *for some* $j \in J_1$, and the $d_{j\ell}$ are given by (7) above. If $\boldsymbol{V}$ is diagonal:

$$q = \sigma_1^2(1-(1/\zeta)) \ and \ s_o^2 \equiv plim\, s^2 = \sigma_1^2/\zeta \ where \ \zeta \equiv (1 + \sum_{j \in J_1} b_{1j}^2(\sigma_1^2/\sigma_j^2)) \tag{16}$$

with $\zeta > 1 \ for \ J_0 \neq \Phi \ and \ \zeta = 1 \ for \ J_0 = \Phi$.

To establish further definitive results, we will assume in the following Propositions 4 - 18 that $\boldsymbol{V}$ is a diagonal matrix, so that the disturbance terms u_{ij} in (4) are contemporaneously uncorrelated across different equations. We then have:

Proposition 4. The cumulative bias, θ_k, is given by:

$$\theta_k \equiv plim\ \hat{\beta}_k - \beta_k = -\frac{\sigma_1^2}{\zeta}\sum_{j\in J_1} b_{1j}(\beta_k b_{1j} + b_{kj})/\sigma_j^2 = -\ s_o^2 \sum_{j\in J_1} b_{1j}(\beta_k b_{1j} + b_{kj})/\sigma_j^2 \qquad (17)$$

for each $k = 2,...,n+m$, and does not depend upon the value of the parameters b_{hj} for any $h \neq 1,k$ for each $j \in J_1$ or upon the value of the parameters $\beta_h \equiv b_{h1}$ for $h \neq 1,k$.

Thus for the case where ***V*** is diagonal, we can establish analytically a smaller set of parameters of the underlying model to which the extent of the cumulative bias is sensitive. For any of the endogenous variables denoted by $k = 2,...,n$, or any of the predetermined variables, as denoted by $k = n+1,....,n+m$, the extent of the cumulative bias θ_k will depend upon the values of the underlying parameters β_k, b_{kj} and b_{1j} for all $j = 2,...,n$, as well as upon the values of σ_j^2 for all $j = 1,...,n$. Again, however, for any given set of endogenous and predetermined variables, the extent of the cumulative bias θ_k does not depend upon the values of the parameters b_{hj} for any other endogenous or predetermined variables (as denoted by $h \neq k,1$) in any of the equations $j = 2,...,n$. Similarly it does not depend upon the values of the coefficients β_h in the first equation on any other endogenous or predetermined variables (as denoted by $h\ (\neq k) = 2,.....,n+m$). Nevertheless, the magnitude of the variances σ_j^2 of the residual disturbance terms will indeed in general vary with which endogenous and predetermined variables are included in these equations.

For each $j \in J_0$, we may define:

$$a_j \equiv b_{1j}^2\sigma_1^2/\sigma_j^2 > 0 \qquad (18)$$

as the ratio between the variance in the jth endogenous variable, x_j, that is due to the impact of the disturbance term u_{i1} on the first variable, x_1, to the variance in x_j that is due to the residual disturbance term in the jth equation, holding constant all other explanatory variables. The following Propositions 5 - 8 follow directly from (16) and (17):

Proposition 5. The cumulative bias, θ_k, will be zero for all $k \in K$ if $J_0 = \Phi$, the null set.

Proposition 6. If $\beta_k \neq 0$, the cumulative proportionate bias θ_k/β_k is equal to a general proportional bias term, given by $\Theta_G \equiv -\sum_{j\in J_0} a_j/\zeta = ((1/\zeta)-1) = (s_o^2/\sigma_1^2)-1$, that is negative whenever $J_0 \neq \Phi$ and equal for all $k \in K$ for which $\beta_k \neq 0$, plus a specific proportional bias term, given by $\Theta_{Sk} \equiv -(\sigma_1^2/\zeta)\sum_{j\in J_0}(b_{1j}b_{kj}/\beta_k\sigma_j^2)$ which varies with the value of each b_{kj}/β_k for $j \in J_0$. If $\beta_k = 0$, the cumulative bias θ_{1k} is equal to simply the specific bias term given by $\Theta'_{Sk} \equiv -(\sigma_1^2/\zeta)\sum_{j\in J_0}(b_{1j}/b_{kj}\sigma_j^2)$.

Proposition 7. It will be sufficient for the cumulative bias, θ_k, to be negative *(positive)* for any $k \in K$ that $J_0 \neq \Phi$ and for each $j \in J_0$ one or more of the following holds: (i) $\beta_k \geq$ *(≤)* 0 and $b_{1j}b_{kj} \geq 0$, (ii) $b_{kj} \geq -\beta_k b_{1j}$ and b_{1j} > *(<)* 0, (iii) $b_{kj} \leq -\beta_k b_{1j}$ and b_{1j} < *(>)* 0, with at least one of the inequalities in (i) - (iii) holding as a strict inequality for at least one $j \in J_0$.

When we define $S_k \equiv \{ j \in J_0 : b_{1j}b_{kj} < 0\}$, $S'_k \equiv \{ j \in J_0 : b_{1j}b_{kj} > 0\}$, we have more generally:

Proposition 8. It is necessary and sufficient for the cumulative bias, θ_k, to be negative *(positive)* for any $k \in K$ that:

$$J_0 \neq \Phi \ \ \& \ \ \sum_{j \in S_k \cup S'_k} b_{kj}(b_{1j} / \sigma_j^2) > (<) - \beta_k \sum_{j \in J_0} (b_{1j}^2 / \sigma_j^2) \tag{19}$$

For a negative cumulative bias, θ_k, when $\beta_k < 0$, Proposition 8 requires that there exist sufficiently large positive products of the coefficients b_{kj} and b_{1j} in some equations $j \in J_0$ to more than offset the absolute values of both any negative products of these coefficients in other equations $j \in J_0$ and the negative values of $\beta_k b_{1j}^2$ for all equations $j \in J_0$, after applying the weights σ_j^{-2}. If the positive products are not sufficiently large to offset the absolute values of these negative terms, Proposition 8 implies that the cumulative bias, θ_k, will be positive whenever $\beta_k < 0$.

Similarly, for a positive cumulative bias, θ_k, when $\beta_k > 0$, Proposition 8 requires there to exist sufficiently large negative products of the coefficients b_{kj} and b_{1j} in some equations $j \in J_0$ to more than offset both any positive products of these coefficients in other equations $j \in J_0$ and the positive values of $\beta_k b_{1j}^2$ for all equations $j \in J_0$, after applying the weights σ_j^{-2}. If the negative products are not large enough to offset them, Proposition 8 implies that the cumulative bias, θ_k, will be negative whenever $\beta_k > 0$.

The *necessary conditions* in Proposition 8 for a positive cumulative bias, θ_k, when $\beta_k > 0$, or for a negative cumulative bias, θ_k, when $\beta_k < 0$, therefore appear to be strong ones. In contrast the *sufficient conditions* in Proposition 7 for a negative cumulative bias, θ_k, when $\beta_k > 0$, or for a positive cumulative bias, θ_k, when $\beta_k < 0$, appear relatively weak. These conclusions are reinforced by Proposition 6, which implies that the cumulative *proportionate* bias θ_k / β_k is equal to a *negative* general term which is the same for all k = 2,...,n + m for which $\beta_k \neq 0$, plus a specific term which will only offset the negative general term if the individual $(b_{1j}b_{kj} / \beta_k \sigma_j^2)$ terms are sufficiently positive overall across the additional inter-relationships j = 2,..., n. If any of the sufficient conditions of Proposition 7 prevails, there will be an *under-estimate* under OLS of the absolute value of the true regression coefficient β_k. This will tend to make it more likely that the regression coefficient β_k will appear to be not significantly different from zero under standard significance tests when it actually has a non-zero value, with an increased risk of an associated *Type II error*. The associated bias in the t-statistic for testing this significance is investigated further in Section 3 below.

In applying Propositions 4 - 8, consideration must be given also to the *stability condition*:

$$\beta_k b_{1k} < 1 \quad \textit{i.e.} \; \beta_k b_{1k} + b_{kk} < 0 \qquad\qquad \text{for all } k \in J_1 \tag{20}$$

that the set of simultaneous equations (4) will involve if they are to yield a stable solution that satisfies the Hicksian stability requirement that B's principal minors of order Λ are positive if Λ is even and negative if Λ is odd (see [7]). (20) implies that for all $k \in J_1$ the combined feedback effect of a unit increase in x_k on x_1 in equation 1 and of the change in x_1 on x_k in equation k is less than the initial unit increase in x_k. If condition (20) does not hold, a series of unstable changes in these variables may prevail. Condition (20) will reinforce any negative cumulative bias θ_k if $b_{1k} < 0$, and condition (20) will reinforce any positive cumulative bias θ_k if $b_{1k} > 0$. We also then have:

Proposition 9. For any $k \in K$, if $J_{0k} \equiv \{j : j \neq k \,\&\, j \in J_0\} \neq \Phi$, b_{1k} < *(>)* 0 when $k \in J_1$, β_k > *(<)* 0, and the stability condition (20) holds, the condition that $b_{1j}b_{kj} \geq (\leq)$ 0 for all $j \in J_{0k}$ is sufficient to ensure that the cumulative bias θ_k is negative *(positive)*. If $\beta_k = 0$, the condition that $b_{1j}b_{kj} \geq (\leq)$ 0 for all $j \in J_0$ and $b_{1j}b_{kj} > (<)$ 0 for some $j \in J_0$ is sufficient to ensure that the cumulative bias θ_k is negative *(positive)*.

Proposition 10. For any $k \in K$, if $J_{0k} \neq \Phi$, $\beta_k > 0 \,\&\, b_{1k} > 0$, or $\beta_k < 0$ & $b_{1k} < 0$, and the stability condition (20) holds, knowledge of the sign pattern of $b_{1j}b_{kj}$ for all $j \in J_{0k}$ alone is insufficient to determine the sign of the cumulative bias θ_k for $k \in J_0$. However, if in addition $|b_{kj}| \geq |\beta_k b_{1j}|$ for all $j \in J_{0k}$, the condition that $b_{1j}b_{kj} \geq$ $(\leq)$ 0 for all $j \in J_{0k}$ is sufficient to ensure that the cumulative bias θ_k is negative *(positive)* whenever b_{1k}< *(>)* 0 and β_k < *(>)* 0 for $k \in J_0$.

For the basic case where n = 2, Proposition 4 implies that:

$$\theta_k = -\sigma_1^2(b_{12}(\beta_k b_{12} + b_{k2}))/(\sigma_2^2 + b_{12}^2\sigma_1^2) \;\text{ for } k = 2,....,2+m \tag{21}$$

and that the simple heuristic $\theta_2 > 0$ iff $b_{12} > 0$, $\theta_2 < 0$ iff $b_{12} < 0$, and $\theta_2 = 0$ iff $b_{12} = 0$ holds under the stability condition (20), irrespective of the value of any other parameters, since $J_{02} = \Phi$, the empty set, in the basic case of n = 2. However, for the predetermined variables $k = 3,....,2+m$ $(m > 0)$, we have $J_{0k} \neq \Phi$ if $b_{12} \neq 0$, even when n = 2. Propositions 4 and 9 then imply that for these predetermined variables, the heuristic becomes: $\theta_k < 0$ if $\beta_k > 0 \,\&\, b_{12}b_{k2} \geq 0$, and $\theta_k > 0$ if $\beta_k < 0 \,\&\, b_{12}b_{k2} \leq 0$, whenever $b_{12} \neq 0$. For any given set of included predetermined variables, the magnitude of the cumulative bias θ_k for any endogenous or predetermined variable k = 2,...., 2 + m, however, does not depend upon the values of the coefficients b_{hj} (for $h \neq 1, k$) on any other (endogenous or predetermined) variables in these two equations.

For the case where n = 3, with now three inter-relationships given by (1) and (2), the set J_{02} is no longer empty. Proposition 9 now implies that for cases where the coefficients $\beta_k \equiv b_{k1}$ and b_{1k}, in the reciprocal relationship between variables k and 1, are *of opposite sign*, we may apply the heuristic:

$$\text{if } \beta_2 > 0, b_{12} < 0, \text{ and } b_{13}b_{23} \geq 0, \text{ then } \theta_2 < 0; \text{ if } \beta_2 < 0, b_{12} > 0, \text{ and } b_{13}b_{23} \leq 0, \text{ then } \theta_2 > 0 \qquad (22)$$

and similarly for the endogenous variable k = 3. For the predetermined variables k = 4,...,2 + m, the heuristic from Proposition 9 becomes:

$$\text{if } \beta_k > 0, \text{ and } b_{1j}b_{kj} \geq 0 \text{ for } j = 2,3, \text{ then } \theta_k < 0; \text{ if } \beta_k < 0, \text{ and } b_{1j}b_{kj} \leq 0 \text{ for } j = 2,3, \text{ then } \theta_k > 0 \quad (23)$$

For cases where the coefficients $\beta_k \equiv b_{k1}$ and b_{1k}, in the reciprocal relationship between variables k and 1, are *of the same sign*, the heuristic (23) still holds for the predetermined variables. However for the endogenous variable k = 2, Proposition 10 implies that the absolute magnitude of b_{23} relative to that of the product of β_2 and b_{13} also now matters, with the extended heuristic becoming:

$$\text{if } \beta_2 < 0, b_{12} < 0, b_{13}b_{23} \geq 0, \text{ and } |b_{23}| \geq |\beta_2 b_{13}|, \text{ then } \theta_2 < 0 \qquad (24)$$

$$\text{if } \beta_2 > 0, b_{12} > 0, b_{13}b_{23} \geq 0, \text{ and } |b_{23}| \geq |\beta_2 b_{13}|, \text{ then } \theta_2 > 0 \qquad (25)$$

Propositions 9 and 10 generalise these heuristic conditions to the more general case of n endogenous variables and $m \geq 0$ predetermined variables. In line with our earlier discussion of the implications of Propositions 6 - 8, the conditions which are sufficient in (24) and (25) to ensure that the absolute value of the true regression coefficient β_2 is less than that of its OLS asymptotic estimate, and hence that $\theta_2 < 0$ when $\beta_2 < 0$, and $\theta_2 > 0$ when $\beta_2 > 0$, are more stringent than those in the heuristic (23). The conditions in the heuristic (23) are themselves sufficient to ensure that the absolute value of the true regression coefficient β_2 is greater than that of its OLS asymptotic estimate, and hence $\theta_2 < 0$ when $\beta_2 > 0$, and $\theta_2 > 0$ when $\beta_2 < 0$.

3. Cumulative Endogeneity Bias and Statistical Inference

In this section, we investigate the implications of our above analysis for our ability to infer restrictions on the true values of the parameters of the primary equation of interest from knowledge of their OLS asymptotic estimates, in the presence of cumulative endogeneity bias. We will examine first the implications for the use of standard t-statistics to test whether or not the true value of a coefficient β_k is significantly different from zero.

Proposition 11. For any $k \in K$, if $J_0 \neq \Phi$ and $\beta_k \neq 0$, the asymptotic proportional bias in the standard OLS t-statistic t_k associated with testing whether the coefficient β_k is significantly different from zero equals:

$$(t_k' - t_{ok}) / t_{ok} = ((1 - (L_k / \beta_k)) / \zeta^{0.5}) - 1 \tag{26}$$

where
$$L_k \equiv \sigma_1^2 \sum_{j \in J_0} (b_{1j} b_{kj} / \sigma_j^2) = \sum_{j \in J_0} a_j (b_{kj} / b_{1j}) \tag{27}$$

and $t_k' \equiv \hat{\beta}_k^o / (s_o \tilde{\varsigma}_{kk}^{0.5}), t_k \equiv \hat{\beta}_k (s\varsigma_{kk}^{0.5}), t_{ok} = \beta_k / (\sigma_1 \tilde{\varsigma}_{kk}^{0.5}),$ for any given value of $\tilde{\varsigma}_{kk} \neq 0,$ where $\tilde{\varsigma}_{kk}$ $\equiv plim\, \varsigma_{kk}$ and ς_{kk} is the (k-1)th element on the principal diagonal of $(\boldsymbol{Y_0}'\boldsymbol{Y_0})^{-1}$. The proportional bias in (26) will be negative whenever $(L_k / \beta_k) \geq 0$, and is a strictly decreasing function of b_{kj} / β_k iff $b_{1j} > 0$, of each b_{1j} iff $(b_{kj} + \hat{\beta}_k^o b_{1j})\beta_k > 0$ and of each a_j in (18) iff $((b_{kj} / b_{1j}) + 0.5\hat{\beta}_k^o)\beta_k > 0$ for each $j \in J_0$. If additionally, $2 > (L_k / \beta_k) \geq 0$, then $|t_k'| < |t_{ok}|$.

Proposition 11 implies that, even with an infinitely large sample of observations, *downward proportionate bias* will result in the standard t-statistic for testing whether the coefficient β_k is significantly different from zero for any endogenous or predetermined variable k = 2,..., n + m, whenever each product $(b_{1j} b_{kj} / \beta_k)$ is non-negative for $j \in J_0$, $\beta_k \neq 0$ and $J_0 \neq \Phi$, for any given value of $\varsigma_{kk} \neq 0$. The downward proportionate bias will be greater the larger is each b_{1j}, b_{kj} and a_j, whenever the corresponding b_{1j}, b_{kj}, β_k and $\hat{\beta}_k^o$ are positive. Since $b_{kk} = -1$ for an endogenous variable $k \in J_1$, a non-negative value of $(b_{1k} b_{kk} / \beta_k)$ implies that β_k and b_{1k} are of opposite sign for such an endogenous variable whenever both are non-zero. If β_k and b_{1k} are of the same sign for an endogenous variable, the sufficient condition $(L_k / \beta_k) \geq 0$ for (26) to be negative implies that not only are the remaining $(b_{1j} b_{kj} / \beta_k)$ terms non-negative for $j \in J_{0k}$ whenever $J_{0k} \neq \Phi$, but more strongly that:

$$\sum_{j \in J_{0k}} (b_{1j} b_{kj} / \beta_k \sigma_j^2) \geq (b_{1k} / \beta_k \sigma_k^2) > 0 \quad \textit{for } k \in J_0 \tag{28}$$

so that again there are more stringent implications for the interaction terms when β_k and b_{1k} are of the same sign for an endogenous variable, than when they are of opposite sign. For a predetermined variable k = n + 1,..., n + m, there is no such restriction that $b_{kk} = -1$, so that there is no such asymmetry when β_k and b_{1k} are of the same sign, in the sufficient conditions for (26) to be negative.

For any endogenous or predetermined variable k, a negative value of (26) when $t_{ok} > 0$ suggests a greater risk of a *Type II error* in a standard one-sided t-test, of accepting the null hypothesis that $\beta_k \leq 0$ when this hypothesis is untrue. Similarly, a negative value of (26) when $t_{ok} < 0$ suggests a greater risk, in a standard one-sided t-test, of the Type II error of accepting the null hypothesis that $\beta_k \geq 0$ when this hypothesis is untrue. If, in addition, $2 > (L_k / \beta_k) \geq 0$, Proposition 11 also implies that under its stated conditions the *absolute value* of the t-statistic t_k' in the presence of endogeneity bias will be less than its absolute asymptotic value in the absence of such bias. This in turn suggests a greater risk in these circumstances of a Type II error under a standard two-sided t-test, of accepting the null hypothesis that $\beta_k = 0$ when this hypothesis is untrue, than would prevail if there had been no such endogeneity bias. However, the validity of relying upon standard t-statistics for making inferences regarding the true value of β_k, even with large samples, is more generally called into question by the existence of endogeneity bias, since the standard proofs of the validity of the associated t-tests (see e.g. [2, 8]) assume a zero contemporaneous correlation between the relevant regressors and the disturbance term in the estimated equation.

We may also demonstrate from equations (16) and (17) that:

$$\hat{\beta}_k^o = (\beta_k - L_k)/\zeta \qquad (29)$$

Hence we have:

Proposition 12. A necessary and sufficient condition for the asymptotic estimate, $\hat{\beta}_k^o$, of β_k to be positive *(negative, zero)* for any $k \in K$ is that β_k is greater than *(less than, equal to)* L_k.

Knowledge of simply the sign of the asymptotic estimate, $\hat{\beta}_k^o$, therefore places an (upper or lower) bound on the value of the true coefficient β_k in terms of the overall interaction effect L_k.

Proposition 13. The true value of the coefficient β_k is related to its OLS asymptotic estimate $\hat{\beta}_k^o$ through the equation:

$$\beta_k = (1+\rho)\hat{\beta}_k^o + L_k \quad \textit{for any } k \in K \textit{ where } \rho \equiv \sum_{j \in J_o} a_j \qquad (30)$$

with $\beta_k > \hat{\beta}_k^0$ whenever $\hat{\beta}_k^o > 0, b_{1j}b_{kj} \geq 0$ for all $j \in J_0$ and $J_0 \neq \Phi$, and $\beta_k < \hat{\beta}_k^o$ whenever $\hat{\beta}_k^o < 0, b_{1j}b_{kj} \leq 0$ for all $j \in J_0$ and $J_0 \neq \Phi$.

In contrast to the potential unreliability of standard t-tests in the presence of endogeneity bias, Proposition 13 above implies that knowledge of the OLS asymptotic estimate $\hat{\beta}_k^o > 0$ will itself be sufficient to ensure that the true value of the corresponding regression coefficient β_k will be such that $\beta_k > \hat{\beta}_k^o > 0$ whenever the qualitative condition $b_{1j}b_{kj} \geq 0$ for all j = 2,...,n (with not all b_{1j} = 0) holds. Similarly, knowledge of the OLS asymptotic estimate $\hat{\beta}_k^o < 0$ will be sufficient to ensure that the true value of the regression coefficient β_k will be such that $\beta_k < \hat{\beta}_k^o < 0$ whenever the qualitative condition $b_{1j}b_{kj} \leq 0$ for all j = 2,...,n (with not all b_{1j} = 0) is satisfied.

The associated heuristic:

$$\text{if } \hat{\beta}_k^o > 0 \text{ and } b_{1j}b_{kj} \geq 0 \text{ for all } j \in J_1 \text{ (with } b_{1j} \neq 0 \text{ for some } j \in J_1\text{), then } \beta_k > \hat{\beta}_k^o > 0 \qquad (31)$$

$$\text{if } \hat{\beta}_k^o < 0 \text{ and } b_{1j}b_{kj} \leq 0 \text{ for all } j \in J_1 \text{ (with } b_{1j} \neq 0 \text{ for some } j \in J_1\text{), then } \beta_k < \hat{\beta}_k^o < 0 \qquad (32)$$

holds from Proposition 13 for all $k \in K \equiv \{2,....,n+m\}$, including both endogenous and predetermined variables. However, in the case of endogenous variables, we again have $b_{kk} = -1$ for k = 2,...,n. (31) and (32) then require the *weakly opposite signs* $b_{1k} \le 0$ if $\hat{\beta}_k^o > 0$ and $b_{1k} \ge 0$ if $\hat{\beta}_k^o < 0$ in order for the heuristic to be applied to an endogenous variable. In the *same sign case* of $b_{1k} > 0$ and $\hat{\beta}_k^o > 0$, where k is an endogenous variable, we may still ensure that $\beta_k > \hat{\beta}_k^o > 0$ when $b_{1j} \ne 0$ for some $j \in J_1$ if there are sufficiently strong positive offsetting values of $b_{1j}b_{kj}$ for $j(\ne k) = 2,...,n$ to ensure $L_k \ge 0$ in (27) and (30). Similarly, in the same sign case of $b_{1k} < 0$ and $\hat{\beta}_k^o < 0$, where k is an endogenous variable, we may still ensure that $\beta_k < \hat{\beta}_k^o < 0$ when $b_{1j} \ne 0$ for some $j \in J_1$ if there are sufficiently strong negative offsetting values of $b_{1j}b_{kj}$ for $j(\ne k) = 2,...,n$ to ensure $L_k \le 0$ in (27) and (30).

We can examine next the benchmark case where the variance in the jth endogenous variable, x_j, that is due to the influence of the disturbance term u_{i1} on the first variable, x_1, is equal to the variance in x_j that is due to the residual disturbance term in the jth equation for all $j \in J_0$, so that we have $a_j = 1$ for all $j \in J_0$ in (18). We then have in equation (30): $\rho = n'$, where n' is the number of equations in J_0, i.e. those for which $b_{1j} \ne 0$. If, in addition, the interaction effect $L_k = 0$, Proposition 13 implies that we have simply:

$$\beta_k = (1+n')\hat{\beta}_k^o \tag{33}$$

so that *the true value of the coefficient* β_k *is here a precise integer multiple of its OLS asymptotic estimate*.

For cases where $\hat{\beta}_k^o(b_{kj}/b_{1j}) \ge 0$ for all $j \in J_0$, we will have instead $L_k \ge 0$ when $\hat{\beta}_k^o > 0$ and $L_k \le 0$ when $\hat{\beta}_k^o < 0$. If, in addition, $a_j \ge 1$ (so that the variance in the jth endogenous variable, x_j, that is due to the influence of the disturbance term u_{i1} on the first variable, x_1, is at least as great as the variance in x_j that is due to the residual disturbance term in the jth equation), with at least one of these inequalities being a strict inequality for some $j \in J_0$, Proposition 13 implies that:

$$\beta_k > (1+n')\hat{\beta}_k^o > 0 \text{ if } \hat{\beta}_k^o > 0 \text{ and } \beta_k < (1+n')\hat{\beta}_k^o < 0 \text{ if } \hat{\beta}_k^o < 0 \tag{34}$$

Thus, if $\hat{\beta}_k^o \ne 0$, the true absolute value of the regression coefficient β_k will here exceed the *product* of the absolute value of its OLS asymptotic estimate and the number of equations (including the first) into which the variable x_1 enters in a non-zero way. Even stronger (upper or lower) bounds on the true value of the regression coefficient as a *multiple* of its OLS asymptotic estimate $\hat{\beta}_k^o$ are thus generated by equation (30) than is implied by use of the heuristic in (31) and (32).

Our ability to place such bounds upon the true value of β_k from Proposition 13 depends *inter alia* upon observing in the OLS multivariate analysis of a sufficiently large sample that $\hat{\beta}_k^o > 0$ when we have *a priori* reasons for believing that the interaction effect $L_k \geq 0$, or upon observing $\hat{\beta}_k^o < 0$ when we have *a priori* reasons for believing that $L_k \leq 0$. If, however, we observe $\hat{\beta}_k^o > 0$ when we have *a priori* reasons for believing that $L_k \leq 0$, or we observe $\hat{\beta}_k^o < 0$ when we have *a priori* reasons for believing that $L_k \geq 0$, the sign of β_k remains indeterminate in Proposition 13. There is then a risk of a *Type I* error, of rejecting the null hypothesis that $\beta_k = 0$ when it is true, if we rely simply upon the magnitude of $\hat{\beta}_k^o$, or its associated t-statistic, for such an inference. Such a risk, moreover, may tend to increase with the absolute value of L_k in (30), since a large, and seemingly (highly) significant, absolute value of the asymptotic estimate $\hat{\beta}_k^o$ will still be consistent with the null hypothesis of $\beta_k = 0$ in (30) if the absolute value of the interaction effect, L_k, is sufficiently large in the opposite direction. Nevertheless, from Proposition 12, we can still infer $\beta_k > L_k$ if $\hat{\beta}_k^o > 0$ and $\beta_k < L_k$ if $\hat{\beta}_k^o < 0$, so that any additional quantitative information on L_k remains of value in assessing the true value of β_k.

4. The Sensitivity of Cumulative Endogeneity Bias

We have identified in Propositions 1 and 4 above the underlying structural parameters of the conceptual model which the extent of the bias θ_k will depend on, and hence be sensitive to. We have also identified those upon which the bias will not be dependent, and therefore not be sensitive to. In this Section, we assess the direction of the associated sensitivity, and the factors which affect the magnitude of the sensitivity, for those parameters to which the bias is sensitive. From (16) and (17) we have:

Proposition 14. For any given $j \in J_0$ and $k \in K$, $\partial\theta_k / \partial\beta_k = (s_o^2 / \sigma_1^2) - 1 < 0$ when $J_0 \neq \Phi$, and $\partial\theta_k / \partial b_{kj} = (\partial\theta_k / \partial\beta_k) / b_{1j}$, so that the numerical value of the cumulative bias θ_k, *ceteris paribus*, decreases with an increase in β_k, and with an increase in b_{kj} whenever $b_{1j} > 0$, but increases with an increase in b_{kj} whenever $b_{1j} < 0$.

Proposition 15. For any given $k \in K$ and $j \in J_1, \partial\theta_k / \partial b_{1j} = -s_o^2(2b_{1j}\hat{\beta}_k^o + b_{kj}) / \sigma_j^2$, so that the numerical value of the cumulative bias θ_k, *ceteris paribus*, increases with a small increase in b_{1j} from an initial value of $b_{1j} = 0$ when $b_{kj} > 0$, and increases with a small decrease in b_{1j} from an initial value of $b_{1j} = 0$ when $b_{kj} > 0$. More generally, for any initial value of $b_{1j}, \partial\theta_k / \partial b_{1j} >(<,=)$ 0 iff $2b_{1j}\hat{\beta}_k^o + b_{kj} <(>,=) 0$.

Thus, while the cumulative bias, θ_k, is indeed sensitive to the value of each b_{1j} parameter, there are still cases in which the local sensitivity $\partial\theta_k / \partial b_{1j}$ of θ_k to small variations in b_{1j} will be zero. It is also of interest to examine how the extent of the cumulative bias varies with the magnitude of the variance of the disturbance term in each equation. From (17), we have:

Proposition 16. For any given $k \in K$, $\partial\theta_k / \partial\sigma_1^2 = \theta_k / \sigma_1^2 \zeta$ where $1/\zeta = s_o^2 / \sigma_1^2 < 1$ for $J_0 \neq \Phi$, so that an increase in the variance of the disturbance term in the first equation results in the same proportionate increase in the cumulative bias θ_k for all endogenous and predetermined variables $k \in K$ for which θ_k is non-zero, whenever the set J_0 is non-empty and hence $b_{1j} \neq 0$ for some $j = 2,...,n$. The local sensitivity $\partial\theta_k / \partial\sigma_j^2$ of θ_k to a small increase in the variance of the disturbance term in the jth equation for any $j \in J_1$ equals $s_o^2 (b_{1j}^2 \hat{\beta}_k^o + b_{1j} b_{kj}) / \sigma_j^4$, and is positive *(negative, zero)* iff $b_{1j}^2 \hat{\beta}_k^o + b_{1j} b_{kj}$ is positive *(negative, zero).*

The quantitative estimates of s_o^2 and $\hat{\beta}_k^o$ that are available from a (large-sample) OLS analysis can therefore be combined with different feasible values of the underlying structural parameters σ_1^2, b_{1j}, b_{kj} and σ_j^2 in Propositions 14 - 16 to form a quantitative assessment of the sensitivity of the magnitude of the cumulative endogeneity bias θ_k to changes in the underlying structural parameters β_k, b_{kj} and b_{1j} to which it is potentially sensitive. Again a quantitative assessment of such sensitivity is an important task for empiricists in the presence of possible endogeneity bias in the available OLS parameter estimates. So too is an awareness of changes in the underlying structural parameters that would leave the extent of the cumulative bias θ_k unchanged, as in Proposition 17 below.

Proposition 17. The cumulative bias θ_k for each $k \in K$ is invariant under changes in (i) the variances of the disturbance terms that leave each ratio σ_j^2 / σ_1^2 unchanged for each $j \in J_1$; (ii) the coefficients b_{1j} and in the variances of the disturbance terms for each $j \in J_1$ that leave the ratios $a_j \equiv b_{1j}^2 \sigma_1^2 / \sigma_j^2$ and b_{kj} / b_{1j} unchanged; and (iii) the sign of b_{1j} that leave its absolute value unchanged, whenever $b_{kj} = 0$ for a given $j \in J_1$.

Propositions 16 and 17 highlight the importance of the impact of the relative disturbance variances σ_j^2 / σ_1^2 upon the size of each cumulative bias. If all of these relative variances do not change, then *ceteris paribus* neither will the extent of each cumulative bias θ_k. However, a larger value of σ_1^2, holding each σ_j^2 constant, will increase the absolute value of each θ_k. On the other hand, a larger value to σ_j^2 for any given $j \in J_1$ will make less negative the extent of the negative cumulative bias θ_k whenever the OLS asymptotic estimate $\hat{\beta}_k^o$ is positive and the sufficient condition from Proposition 9, that $b_{1k} > 0, \beta_k > 0, b_{1j} b_{kj} \geq 0$ for all $j \in J_{0k}$, for θ_k to be negative under the stability condition (20), holds. A greater variance in the disturbance term in the jth inter-relationship, for j = 2,...,n, will then be positively beneficial in reducing the absolute magnitude of the cumulative bias. Such an increased variance, holding that for equation one constant, will map out a more extensive set of intersection points with equation one and the other inter-relationships in (3) that more accurately traces out the slope parameter β_k in equation one when an OLS multivariate regression plane is fitted to the resultant intersection points.

The sensitivity of the cumulative bias to a change in the specification of the underlying conceptual model through the inclusion or deletion of any given equation $j \in J_1$ within the model, holding all other parameters constant, is examined in the following Proposition, where θ_k denotes the extent of the cumulative bias when equation $r \in J_1$ is included in the model, and θ_k^r its numerical value when equation r is not included in the model, with all other equations in the model and their parameters remaining unchanged.

Proposition 18. For any $r \in J_1$ and any $k \in K$, $\theta_k - \theta_k^r$ is positive *(negative, zero)* iff $b_{1r}^2 \hat{\beta}_k^o + b_{1r} b_{kr}$ is negative *(positive, zero)*, with $b_{1r} = 0$ being sufficient for $\theta_k = \theta_k^r$.

Thus if the asymptotic OLS estimate, $\hat{\beta}_k^o$, is positive, and b_{1r} is non-zero and weakly of the opposite sign to b_{kr}, we will have $b_{1r}^2 \hat{\beta}_k^o + b_{1r} b_{kr}$ positive. Proposition 18 then implies that, if the bias is negative in the absence of the rth interrelationship, the existence of the rth interrelationship will make the extent of the negative cumulative bias even greater.

5. Partial Regressions and the Impact of Instruments

The cumulative endogeneity bias that we have analysed so far arises when an OLS multivariate regression is carried out using all the endogenous and predetermined variables as explanatory variables. The extent of the cumulative endogeneity bias in the estimated coefficient for any given included variable will, however, in general depend upon which set of variables is used in the regression. One reason to exclude a variable from the OLS multivariate regression is a lack of data on this variable. Another is a belief concerning its apparent lack of statistical analysis, for which the available t-statistics may not provide a reliable guide, as we have discussed in Section 3 above. In this section, we therefore extend our analysis to consider the impact on the extent of the cumulative bias of excluding some endogenous and/or predetermined variables from the OLS multivariate regression. We will examine first the case where only predetermined variables are excluded from the OLS regression for the primary equation of interest.

Proposition 19. If the true underlying structural model is that given by equations (1) - (2), but only the predetermined variables z_h for $h \in M' \equiv \{1,...,m'\}$ where $m' < m$, together with the endogenous variables x_k for $k \in J_1 \equiv \{2,...,n\}$, are included as regressors in the OLS multivariate regression analysis of equation (1), the resulting overall cumulative bias $\theta_k'' \equiv plim\hat{\beta}_k - \beta_k$ for each for $k \in K' \equiv \{2,...,n+m'\}$ equals:

$$\theta_k'' = \theta_k^o + \theta_k' = \theta_k + \sum_{h \in M''} \tau_{kh}^* \hat{\beta}_{n+h}^o \quad \text{where} \quad \theta_k^o = \sum_{h \in M''} \tau_{kh}^* \beta_{n+h}, \; \theta_k' = \theta_k + \sum_{h \in M''} \tau_{kh}^* \theta_{n+h} \qquad (35)$$

and where $M'' \equiv \{m'+1,...,m\}$. The weights $[\tau_{kh}^*] \equiv \boldsymbol{T}^*$ are the asymptotic values of the OLS regression coefficients in the regression of each excluded predetermined variable z_h for $h \in M''$ on the set of included variables x_k for $k \in K'$. For each $k \in K'$, we have also:

$$\theta_k'' = -[\sum_{j \in J_1} (b_{kj} + \beta_k b_{1j}) \sum_{\ell \in J_1} d_{j\ell}^o (b_{1\ell} v_{11} + v_{1\ell})] \;\text{ for }\; [d_{j\ell}^o] \equiv [b_{1j}(b_{1\ell} v_{11} + v_{1\ell}) + v_{j1} b_{1\ell} + v_{j\ell}]^{-1} \; j, \ell \in J_1 \qquad (36)$$

where $[v_{j\ell}] \equiv \boldsymbol{V_o} \equiv \boldsymbol{V} + \boldsymbol{C_o}'\boldsymbol{\Omega_o C_o}$ for $j, \ell \in J$, $\boldsymbol{C_o} \equiv [c_{hj}]$ *for* $h \in M''$, $j \in J$, $\boldsymbol{\Omega_o} \equiv [\varpi_{h\ell}]$ for $h, \ell \in M''$ is the $(m-m')$ x $(m-m')$ (positive definite) covariance matrix for the $m-m'$ excluded predetermined variables, and $\boldsymbol{V}$ is not necessarily diagonal. For the case where $\boldsymbol{V_o}$ is diagonal, the cumulative bias also equals:

$$\theta_k'' = -v_{11} \sum_{j \in J_o} (b_{1j}(\beta_k b_{1j} + b_{kj}) / v_{jj}\xi) \text{ where } \xi \equiv (1 + v_{11} \sum_{j \in J_1} (b_{1j}^2 / v_{jj})) > 0 \tag{37}$$

Proposition 19 implies that the overall cumulative bias, θ_k'' , for any included endogenous or predetermined variable $k \in K'$, when $m-m'$ predetermined variables are excluded from the OLS regression, is decomposable into two parts, as in equation (35). The first part, θ_k^o , is associated with omitted variables bias [2], and equals zero if the coefficients $\beta_{n+h} = c_{h1}$ on the excluded predetermined variables in the first equation are zero. The second part, θ_k' , is associated with the simultaneity bias that arises from the multiple relationships that relate the regressor to disturbance term in the first equation. As in (35), θ_k' equals the magnitude of the cumulative endogeneity bias θ_k for $k \in K'$ when all of the predetermined and endogenous variables are included in the OLS regression, plus a weighted sum of the magnitudes of the cumulative endogeneity biases θ_{n+h} for the excluded predetermined variables when all of the predetermined and endogenous variables are included in the OLS regression. As in equation (35), the overall bias when some predetermined variables are excluded from the OLS estimation of the first equation also equals the cumulative endogenous bias when they are included in the OLS estimation, plus a weighted sum of the OLS asymptotic parameter estimates for the excluded variables that are generated when all of the predetermined variables are included in the OLS estimation. In each case, the weights are the asymptotic regression coefficients that result from OLS multivariate regressions of each excluded predetermined variable on the set of included variables.

Parallel conditions to those of Proposition 2 hold for determining the sign of the overall cumulative bias, θ_k'' , in (35) for any included variable $k \in K'$, but with each σ_{kj} replaced by v_{kj} for all $k, j \in J$. Similarly, if $\boldsymbol{V_o}$ is diagonal, parallel forms of Propositions 5 – 18 hold for the overall cumulative bias, θ_k'' , in (37) when $m-m'$ predetermined variables are excluded from the OLS regression as hold for the cumulative bias, θ_k , when they are not excluded. However, the condition that $\boldsymbol{V_o}$ is diagonal is now a stronger one than the earlier condition that $\boldsymbol{V}$ is diagonal. Conditions under which $\boldsymbol{V_o}$ will be diagonal are that (i) $\boldsymbol{V}$ is diagonal, (ii) any given excluded predetermined variable enters into no more than one of the equations $j \in J$, so that for all $h \in M''$: if $c_{hj'} \neq 0$ for some $j' \in J$, then $c_{hj} = 0$ for all $j \in J$, and (iii) excluded predetermined variables that appear in different structural equations are uncorrelated, so that there is a zero covariance $\varpi_{h\ell} = 0$ when $c_{hk} \neq 0$ and $c_{\ell j} \neq 0$ for $k \neq j \,\&\, k, j \in J$. Excluding predetermined variables that enter into the first equation in a non-zero way will then have the effect of making the associated variance v_{11} of the resulting residual disturbance term of the first equation greater in (37) than the original variance σ_1^2 of the disturbance term for the first equation in (17), given the positive definitiveness of the covariance matrix $\boldsymbol{\Omega_o}$ in (36). As in Proposition 16, this will in turn result in a proportionate increase in the extent of the cumulative bias, if it is initially non-zero.

Excluding predetermined variables that enter into another equation $j > 1$ in a non-zero way will, however, have the effect of making v_{jj} in (37) greater than σ_j^2 in (17), and be equivalent to an increase in the variance of the disturbance term of the jth equation. As in Proposition 16, other things being equal, this will make any initial negative cumulative bias in estimating β_k for any included variable *less negative*, so long as $b_{1j}^2\hat{\beta}_k^o + b_{1j}b_{kj} > 0$ throughout this change. If this condition holds and $\boldsymbol{V_o}$ is diagonal, there is scope for *reducing* the extent of any initial negative cumulative bias without increasing the disturbance term for equation one, by excluding predetermined variables z_h from the OLS regression whenever we can impose the exclusion restriction $c_{h1} = 0$ when $c_{hj} \neq 0$ for some $h \in M$ and $j \in J_1$. This exclusion restriction will in particular hold if identifiability of the first equation is secured through imposing restrictions upon the structural parameters in $\boldsymbol{B}$ and $\boldsymbol{C}$, since this involves the associated necessary order condition [8, p. 455] that the number of predetermined variables which are excluded from the first equation must be at least as great as the number of endogenous variables included in it less one. Excluding these predetermined variables from the OLS regression reduces the cumulative bias by making the effective variances v_{jj} in (37) of the disturbance terms inclusive of the influence of these predetermined variables in the equations in which they do appear greater than the corresponding variances σ_j^2 in (17), without at the same time making v_{11} greater than σ_1^2 in the first equation.

We will examine next the case where some *endogenous* variables, x_k for $k \in J'' \equiv \{n'+1,...,n\}$, as well as some predetermined variables, z_h for $h \in M'' \equiv \{m'+1,...,m\}$, are excluded from the OLS multivariate regression for the first equation. We can partition the matrices $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{V}$ as:

$$\boldsymbol{B} \equiv \begin{pmatrix} \boldsymbol{B_{11}} & \boldsymbol{B_{12}} \\ \boldsymbol{B_{21}} & \boldsymbol{B_{22}} \end{pmatrix},\ \boldsymbol{C} \equiv \begin{pmatrix} \boldsymbol{C_{11}} & \boldsymbol{C_{12}} \\ \boldsymbol{C_{21}} & \boldsymbol{C_{22}} \end{pmatrix},\ \boldsymbol{V} \equiv \begin{pmatrix} \boldsymbol{V_{11}} & \boldsymbol{V_{12}} \\ \boldsymbol{V_{21}} & \boldsymbol{V_{22}} \end{pmatrix} \tag{38}$$

where $\boldsymbol{B_{11}} \equiv [b_{kj}]$ for $k, j \in J' \equiv \{1,...,n'\}$, $\boldsymbol{C_{11}} \equiv [c_{hj}]$ for $h \in M', j \in J'$, $\boldsymbol{V_{11}} \equiv [\sigma_{kj}]$ for $k, j \in J'$, and $\boldsymbol{B_{22}}$ is assumed to be non-singular. We can then derive:

Proposition 20. If the true underlying structural model is that given by equations (1) - (2), but only the endogenous variables x_k for $k \in J_1' \equiv \{2,...,n'\}$ where $n' < n$, and the predetermined variables z_h for $h = 1,..,m' < m$, are included as regressors in the OLS multivariate regression analysis of equation (1), the resulting overall cumulative bias $\theta_k''' \equiv plim\hat{\beta}_k - \beta_k$ for each included variable $k \in K'' \equiv \{2,...,n', n+1,...,n+m'\}$ equals:

$$\theta_k''' = -[\sum_{j \in J_1'}(b_{kj}'' + b_{k1}''b_{1j}'')\sum_{\ell \in J_1'} d_{j\ell}''(b_{1\ell}''v_{11}'' + v_{1\ell}'')] \text{ for } [d_{j\ell}''] \equiv [b_{1j}''(b_{1\ell}''v_{11}'' + v_{1\ell}'') + v_{j1}b_{1\ell}'' + v_{j\ell}'']^{-1}\ j, \ell \in J'' \tag{39}$$

where $[b_{kj}''] \equiv (\boldsymbol{B_{11}} - \boldsymbol{B_{12}A_1})\boldsymbol{A_o}$ for $k, j \in J'$, $[b_{n+k,j}''] \equiv (\boldsymbol{C_{11}} - \boldsymbol{C_{12}A_1})\boldsymbol{A_o}$ for $k \in M', j \in J'$ and $\boldsymbol{V''} \equiv$

$[v''_{kj}] \equiv \boldsymbol{A}'_o(\boldsymbol{V}_{11} - \boldsymbol{V}_{12}\boldsymbol{A}_1 - \boldsymbol{A}'_1\boldsymbol{V}_{21} + \boldsymbol{A}'_1\boldsymbol{V}_{22}\boldsymbol{A}_1 + \boldsymbol{A}'_2\boldsymbol{\Omega}_o\boldsymbol{A}_2)\boldsymbol{A}_o$ for $k, j \in J'$, $\boldsymbol{A}_o \equiv [\delta_{kj} / \vartheta_j]$ for $k, j \in J'$, $\boldsymbol{A}_1 \equiv \boldsymbol{B}_{22}^{-1}\boldsymbol{B}_{21}$, $\boldsymbol{A}_2 \equiv \boldsymbol{C}_{21} - \boldsymbol{C}_{22}\boldsymbol{A}_1$, δ_{kj} is Kronecker's delta and ϑ_j is the element in the jth row and jth column of the matrix $(\boldsymbol{B}_{12}\boldsymbol{A}_1 - \boldsymbol{B}_{11})$. If $\boldsymbol{V}''$ is diagonal, we have also:

$$\theta'''_k = -v''_{11} \sum_{j \in J'_1} (b''_{1j}(b''_{k1}b''_{1j} + b''_{kj}) / v''_{jj}\xi'') \text{ where } \xi'' \equiv (1 + v''_{11} \sum_{j \in J'_1} ((b''_{1j})^2 / v''_{jj})) > 0 \tag{40}$$

In addition to the general case considered by Proposition 20, we can examine here the particular case where we exclude only a single endogenous variable, x_n, together with the omitted predetermined variables z_h for $h = m' + 1, ..., m$, so that the OLS regression model becomes:

$$x_{i1} = \sum_{k=2}^{n-1} \beta_k x_{ik} + \sum_{h=1}^{m'} \beta_{n+h} z_{ih} - u''_{i1} \tag{41}$$

where u''_{i1} is the new associated disturbance term for the first equation. If the true underlying structural equations (1) - (2) still apply, we have from Proposition 20:

$$v''_{kj} = ((\sigma_{kj} + \sigma_{kn}b_{nj} + b_{nk}\sigma_{nj} + b_{nk}\sigma_{nn}b_{nj}) / \vartheta_k\vartheta_j) + \sum_{\ell \in M''} \sum_{h \in M''} b''_{n+h,k}\varpi_{h\ell}b''_{n+\ell,j} \quad \textit{for } k, j \in J' \tag{42}$$

$$b''_{kj} = (b_{kj} + b_{kn}b_{nj}) / \vartheta_j \;\; \textit{and} \;\; \vartheta_j = 1 - b_{jn}b_{nj} \;\; \textit{for } k \in K''' \equiv \{1, ..., n', n+1, ..., n+m'\} \;\&\; j \in J' \tag{43}$$

with $\boldsymbol{B}_{22} = b_{nn} = -1$ if we exclude only the single endogenous variable, x_n, and $\vartheta_j > 0$ under similar stability conditions to (20) for each $j \in J'$.

The coefficient b''_{kj} in (43) reflects *the net effect* of variable k on x_j taking account of not only the influence of variable k on x_j in equation j, but also the influence of variable k in equation n upon the excluded endogenous variable, x_n, and the associated influence of x_n upon x_j in equation j of the underlying structural model. If more than one endogenous variable is excluded from the OLS regression, a similar net effect is involved for each coefficient in $[b''_{kj}] \equiv (\boldsymbol{B}_{11} - \boldsymbol{B}_{12}\boldsymbol{A}_1)\boldsymbol{A}_o$ in (39). If there is sufficient *a priori* or other evidence to determine the sign pattern of these net effects, then parallel propositions to Proposition 2 and 5 – 18 may be derived for the sign and sensitivity of the overall cumulative bias given by equations (39) and (40) of Proposition 20 in terms of these net effects.

However, when we exclude one or more endogenous variables from the OLS multivariate regression, the conditions for the associated covariance matrix $\boldsymbol{V}''$ to be diagonal become more stringent than the conditions for $\boldsymbol{V}_o$ and $\boldsymbol{V}$ to be diagonal. In the case of a single excluded endogenous variable, it will nevertheless be sufficient for $\boldsymbol{V}''$ to be diagonal that (i) $\boldsymbol{V}$ is diagonal, (ii) that any given excluded predetermined variable enters in a non-zero way into no more than one of the equations j= 1,…, n, (iii) that excluded predetermined variables which appear in different structural equations are uncorrelated, and (iv) that the excluded endogenous variable enters in a non-zero way into no more than one of the equations j = 1,…, n - 1. If $\boldsymbol{V}''$ is diagonal, and we have knowledge of the sign pattern of the relevant $b''_{\ell j}$ net effect parameters, we can apply parallel versions of Propositions 5 - 18 to an assessment of the overall cumulative bias θ''_k.

By way of illustration, we can examine the case where we exclude a single endogenous variable x_n from the OLS regression, conditions (i) – (iv) for $\boldsymbol{V}''$ to be diagonal hold, and x_n only enters in a non-zero way into equations 1 and n. Since we then have $b_{nj} = 0$ for $j = 2,...,n-1$, (43) implies that $b''_{kj} = b_{kj}$ for $j = 2,...,n-1$, for all the included endogenous and predetermined variables $k \in K''$ and for $k = 1$. Knowledge of the sign pattern of these b_{kj} therefore implies knowledge here of the sign pattern of the corresponding net effect b''_{kj}. Proposition 9 then provides sufficient conditions for determining the sign of the cumulative bias for any given included endogenous and predetermined variables, but now applied to equations $j = 2,...,n-1$ and to the stability condition $b''_{k1}b_{1k} < 1$ for $k = 2,...,n-1$. Thus if n = 4, and we exclude the variable x_4 from the OLS regression for the first equation because of lack of data or other reasons, we can still here sign the overall cumulative biases θ'''_2 and θ'''_3, as well as those for the coefficients of the predetermined variables that are included in the first equation, when these sufficient conditions hold.

If $\boldsymbol{V}''$ is not diagonal, we can no longer rely on parallel versions of Propositions 5 – 18 to derive the sign and sensitivity of the overall cumulative bias. Instead, we must satisfy the more stringent necessary and sufficient conditions of a parallel version of Proposition 2, as applied to the parameters of (39) rather than to those of equations (6) and (7), to be able to determine the qualitative sign of the overall cumulative bias θ'''_k from qualitative information on the underlying structural parameters and the components of $\boldsymbol{V}''$. If these necessary and sufficient conditions are not satisfied, equation (39), like equations (6) and (7), nevertheless provides a means of generating the numerical probability distribution of the extent of the overall cumulative bias through a process of Monte Carlo numerical simulation ([10]), if there is sufficient information upon which to base an assessment of the probability distribution of the underlying parameters that enter into these equations.

However, even when $\boldsymbol{V}''$ is diagonal, we can show that replacing an endogenous variable, such as x_n, by the instrument of a proxy variable that is correlated with x_n but uncorrelated with the disturbance term u_{i1} in the first equation, will not necessarily reduce the absolute magnitude of the cumulative bias, but may instead *increase* it. One such proxy variable is provided by the predetermined variable z_1 if the nth equation is of the form:

$$x_{in} = b_{1n}x_{i1} + b_{n+1,n}z_{i1} - u_{in} \tag{44}$$

where b_{1n} and $b_{n+1,n}$ are non-zero, and z_1 only enters into the nth equation. We can then compare the extent of the cumulative bias θ'''_k when x_n is replaced by its instrument z_1 with the extent of the bias θ''_k that prevails when x_n is included as a regressor for the first equation and only its proxy z_1 is excluded. Under the assumption that $\boldsymbol{V}$, and hence here $\boldsymbol{V_o}$, is diagonal, the extent of the bias θ''_k is given by equation (37), with $v_{jj} = \sigma_j^2$ for n = 1,..., n – 1 and $v_{nn} = \sigma_n^2 + b_{n+1,n}^2\varpi_{11}$ in this case. Under the assumption also that $b_{1k} < 0, \beta_k > 0, b_{1j}b_{kj} \geq 0$ for all $j(\neq k) = 2,...,n$, and that the stability condition (20) holds, we have from Proposition 9 that $\theta_k'' < 0$.

When we simply replace the endogenous variable x_n by its proxy z_1 as a regressor for the first equation, the assumption that $\boldsymbol{V}$ is diagonal implies here that $\boldsymbol{V}''$ is also diagonal, with the extent of the resulting cumulative bias θ_k''' given by equations (40), (42) and (43). Under the above assumptions, we now have $v_{11}'' = \sigma_1^2 + b_{n1}^2\sigma_n^2$, $v_{jj}'' = \sigma_j^2$ for $j = 2,...,n-1, b_{kn} = 0 = b_{jn}, \vartheta_j = 1$ and hence $b_{kj}'' = b_{kj}$ for $j = 2,...,n-1$ and $k \in K_o \equiv \{2,...,n-1, n+2,...,n+m\}$. For these endogenous and predetermined variables $k \in K_o$, we therefore have from (37) and (40):

$$\theta_k'' = -(\sigma_1^2/\xi)((\sum_{j=2}^{n-1} b_{1j}(\beta_k b_{1j} + b_{kj})/\sigma_j^2) + \lambda_k) \textit{ where } \lambda_k \equiv b_{1n}^2\beta_k/(\sigma_n^2 + b_{n+1,n}^2\varpi_{11}) \tag{45}$$

$$\theta_k''' = -(v_{11}''/\xi'')(\sum_{j=2}^{n-1} b_{1j}''((\beta_k/\vartheta_1)b_{1j}'' + b_{kj}))/\sigma_j^2) \textit{ where } b_{1j}'' = b_{1j} + b_{1n}b_{nj} \tag{46}$$

The substitution of the instrument z_1 for the endogenous variable x_n frees (46) from the positive term λ_k in (45) that in itself tends to make the absolute value of θ_k''' less than that of θ_k''. However, if $b_{n1} \neq 0$, the substitution of x_n by its imperfect proxy z_1 increases the variance of the disturbance term of the associated OLS regression equation, with the result that $(v_{11}''/\xi'') >$ (σ_1^2/ξ). If $|b_{1j}''| > |b_{1j}|$ for all j=2,..., n-1 and $b_{1n}b_{n1} \geq 0$, so that $0 < \vartheta_1 \leq 1$, the absolute magnitude of θ_k''' will exceed that of θ_k'', despite the positive term λ_k in (45), if $|b_{nj}|$ is sufficiently large for each j = 2,…, n - 1. If we choose units for the instrument z_1 so that $b_{n+1,n} = 1$ in equation (44), we can also compare the absolute magnitude of the cumulative bias θ_n'' in the OLS asymptotic estimate of β_n when the endogenous variable x_n is included as a regressor, with the bias:

$$\theta_{n+1}^o \equiv \hat{\beta}_{n+1}^o - \beta_n = \theta_{n+1}''' - \beta_n \textit{ since } \beta_{n+1} = b_{n+1,1} = 0 \tag{47}$$

associated with the OLS asymptotic estimate $\hat{\beta}_{n+1}^o$ of the coefficient of z_1 when the instrument z_1 is substituted for x_n in the multivariate regression. From (37) and (40), we have:

$$\theta_n'' = -(\sigma_1^2/\xi)((\sum_{j=2}^{n-1} b_{1j}(\beta_n b_{1j} + b_{nj})/\sigma_j^2) + \lambda_n) \textit{ where } \lambda_n \equiv b_{1n}(\beta_n b_{1n} - 1)/(\sigma_n^2 + \varpi_{11}) \tag{48}$$

$$\theta_{n+1}^o = -(v_{11}''/\xi'')(\sum_{j=2}^{n-1} b_{1j}''((\beta_n/\vartheta_1)b_{1j}'' + b_{nj}))/\sigma_j^2) - \beta_n \textit{ where } b_{1j}'' = b_{1j} + b_{1n}b_{nj} \tag{49}$$

Again we have $(v_{11}''/\xi'') > (\sigma_1^2/\xi)$, so that if $|b_{1j}''| \geq |b_{1j}|$ for all $j = 2,...,n-1$ and $b_{1n}b_{n1} \geq 0$, the absolute magnitude of θ_{n+1}^o will exceed that of θ_n'', whenever $\beta_n \geq (\sigma_1^2/\xi)\lambda_n$.

Thus even if we can find a proxy variable, such as z_1, that is correlated with the endogenous variable x_n and uncorrelated with the disturbance term u_{i1} in the primary equation of interest, its substitution as an instrument for the endogenous variable x_n on a piecemeal basis when n > 2 may not reduce the absolute magnitude of the cumulative bias, but instead may *increase* it, not only for the coefficient β_n, but also for the coefficients of all the other endogenous and predetermined variables.

As in Proposition 16, unless all endogenous variables have been effectively eliminated, so that the set J_0 is empty, the absolute magnitude of any initial cumulative bias for any coefficient β_k increases *ceteris paribus* with the variance of the disturbance term in the first equation. The substitution of an imperfect proxy variable for x_n in effect increases this variance, and thereby risks *increasing* the absolute magnitude of the cumulative bias for all of the coefficients β_k, if there still remain some endogenous variables, and even though the instruments that are used are uncorrelated with the original disturbance term in the first equation. If the instruments used are correlated with this disturbance term, as in Nakamura and Nakamura [12], there is an additional source of risk that their use may increase the magnitude of the endogeneity bias. The use of Instrumental Variable (IV) estimation, in contrast, depends upon finding a whole set of n - 1 valid instruments to substitute for all of the n – 1 endogenous variables, to secure the elimination of the cumulative endogeneity bias for any given coefficient, such as β_n. The task of finding such a complete set of instruments is typically more difficult than finding a single instrument for any given endogenous variable that may be of prime policy or decision-making interest. Moreover, even if such a complete set of instruments is available, IV estimators will in general still remain biased in finite samples (see [8, p. 365]).

6. Conclusion

If a complete set of valid instruments is not available for all of the endogenous variables, the use of instruments as proxy variables for a subset of the endogenous variables may result in an *increase* in the absolute magnitude of the cumulative bias in each of the estimated regression coefficients. In the absence of a complete set of valid instruments, it becomes important to understand more fully the underlying factors to which the direction and extent of cumulative bias in each of these coefficients are sensitive, and those to which they are not sensitive. In highlighting these factors, we have also established conditions under which it is possible to predict the sign of the cumulative bias associated with OLS parameter estimates, and to place upper or lower bounds upon the true values of the underlying parameters based upon the OLS asymptotic estimates.

In interpreting the results of OLS multivariate analyses, such as in meta-studies (e.g. [4, 5]) of the many existing empirical studies that have deployed OLS to estimate parameters of policy or decision-making importance, there is a need to go beyond an examination of the primary equation of interest, to consider wider evidence on the sign and magnitude of those factors to which the extent of cumulative bias is sensitive, and which enter into other relevant interrelationships. Armed with such wider evidence, there is scope for progress to be made in assessing the direction and extent of the cumulative bias, and hence in drawing conclusions on the true values of the key parameters of interest.

Appendix

Proof of Proposition 1. Using (4), we may write:

$$\boldsymbol{X} = [\boldsymbol{x}, \boldsymbol{X_0}] \text{ where } \boldsymbol{x} \equiv [\mathrm{x}_{i1}], \quad \boldsymbol{X_0} \equiv [\mathrm{x}_{ik}] \text{ for i =1,..., p; k = 2,..., n} \tag{A.1}$$

$$\boldsymbol{Y} \equiv [\boldsymbol{X}, \ \boldsymbol{Z}] \equiv [\mathrm{y}_{ik}] \text{ for i=1,..., p; k = 1,...,n + m, } \boldsymbol{Y_0} \equiv [\mathbf{X}_0, \boldsymbol{Z}] \tag{A.2}$$

Using (3), the first structural equation in (4) is of the form:

$$\boldsymbol{x} = \boldsymbol{Y_0}\boldsymbol{\beta} + \boldsymbol{\upsilon} \tag{A.3}$$

where $\boldsymbol{\beta} \equiv [\beta_k]$ for k = 2, ..., n + m and $\boldsymbol{\upsilon} \equiv [-u_{i1}]$ for i =1,..., p. From [8], we have the OLS estimator of the coefficients of (A.3) given by:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{Y_0'Y}_0)^{-1}\boldsymbol{Y}_0'\mathbf{x} \ = (\boldsymbol{Y_0'Y}_0)^{-1}\boldsymbol{Y_0'Y}_0\boldsymbol{\beta} \ + \ (\boldsymbol{Y_0'Y}_0)^{-1}\boldsymbol{Y_0'\upsilon} \ = \ \boldsymbol{\beta} + (\boldsymbol{Y_0'Y}_0)^{-1}\boldsymbol{Y_0'\upsilon} \tag{A.4}$$

with the asymptotic bias given by:

$$\boldsymbol{\theta} = [\theta_k] \equiv \mathit{plim}\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \ = \mathit{plim}\ (\boldsymbol{Y_0'Y_0})^{-1}\boldsymbol{Y_0'\upsilon} \tag{A.5}$$

Using (A.2), we may write:

$$\boldsymbol{Y_0'Y_0} = \begin{pmatrix} \boldsymbol{X_0'X_0} & \boldsymbol{X_0'Z} \\ \boldsymbol{Z'X_0} & \boldsymbol{Z'Z} \end{pmatrix} \text{ with } (\boldsymbol{Y_0'Y_0})^{-1} \equiv \begin{pmatrix} \boldsymbol{P} & \boldsymbol{Q} \\ \boldsymbol{R} & \boldsymbol{S} \end{pmatrix} \tag{A.6}$$

where $\boldsymbol{P}$ is (n-1) x (n-1), $\boldsymbol{Q}$ is (n -1) x m, $\boldsymbol{R}$ is m x (n –1), and $\boldsymbol{S}$ is m x m. From [3, p. 109], we have:

$$\boldsymbol{P} = \boldsymbol{H}^{-1} \ \mathit{where}\ \boldsymbol{H} \equiv \boldsymbol{X_0'X_0} - \boldsymbol{X_0'Z(Z'Z)}^{-1}\boldsymbol{Z'X_0}, \ \boldsymbol{Q} = -\boldsymbol{PX_0'Z(Z'Z)}^{-1}, \boldsymbol{R} = -(\boldsymbol{Z'Z})^{-1}\boldsymbol{Z'X_0P} \tag{A.7}$$

$$\boldsymbol{S} = (\boldsymbol{Z'Z})^{-1} - (\boldsymbol{Z'Z})^{-1}\boldsymbol{Z'X_0Q}, \ (\boldsymbol{X_o'X_o})^{-1} = \boldsymbol{P} - \boldsymbol{QS}^{-1}\boldsymbol{R}, (\boldsymbol{X_o'X_o})^{-1}\boldsymbol{X'Z} = -\boldsymbol{QS}^{-1} \tag{A.8}$$

Using (3) and (4), we may write:

$$\boldsymbol{X} = \boldsymbol{UB}^{-1} - \boldsymbol{ZCB}^{-1} \ \mathit{with}\ \boldsymbol{B} = \begin{pmatrix} -1 & \boldsymbol{\alpha} \\ \boldsymbol{\varphi} & \boldsymbol{B_0} \end{pmatrix} \ \mathit{and}\ \boldsymbol{B}^{-1} = \begin{pmatrix} \boldsymbol{\mu} & \boldsymbol{\gamma} \\ \boldsymbol{\phi} & \boldsymbol{F} \end{pmatrix} \tag{A.9}$$

where $\boldsymbol{\alpha} \equiv [b_{1j}]$ for $j = 2,...,n$, $\boldsymbol{\varphi} \equiv [b_{k1}] = [\beta_k]$ for $k = 2,...,n$, $\boldsymbol{B_0} \equiv [b_{kj}]$ for $k, j = 2,...,n$, and where $\boldsymbol{\mu}$ is 1x1, $\boldsymbol{\gamma}$ is 1x(n – 1), $\boldsymbol{\phi}$ is (n-1)x1, and $\boldsymbol{F}$ is (n – 1)x(n – 1). From [3, p. 109]:

$$\boldsymbol{F} = (\boldsymbol{B_0} + \boldsymbol{\varphi\alpha})^{-1}, \ \boldsymbol{\phi} = \boldsymbol{F\varphi}, \ \boldsymbol{\gamma} = \alpha\boldsymbol{F}, \ \boldsymbol{\mu} = \boldsymbol{\alpha\phi} - 1 \tag{A.10}$$

Let $$\boldsymbol{G}' \equiv [\boldsymbol{\gamma}', \boldsymbol{F}'] = \boldsymbol{F}'[\boldsymbol{\alpha}', \boldsymbol{I}] \tag{A.11}$$

using (A.10).Then from (A.1), (A.9) and (A.11):

$$X_0 = UG - ZCG,\; X_0'Z = G'U'Z - G'C'Z'Z,\; X_0'Z(Z'Z)^{-1} = G'C'Z'Z(Z'Z)^{-1} - G'C' \tag{A.12}$$

$$Z'X_0 = Z'UG - Z'ZCG,\; X_0'X_0 = G'U'UG - G'C'Z'UG - G'U'ZCG + G'C'Z'ZCG \tag{A.13}$$

$$X_0'Z(Z'Z)^{-1}Z'X_0 = G'U'Z(Z'Z)^{-1}Z'UG - G'U'ZCG - G'C'Z'UG' + G'C'Z'ZCG \tag{A.14}$$

Hence from (A.7), (A.12) – (A.14):

$$H = G'U'UG - G'U'Z(Z'Z)^{-1}Z'UG \tag{A.15}$$

Since the disturbance terms and the predetermined variables in (4) are uncorrelated, we have:

$$plim(p^{-1}U'Z) = 0,\; plim(p^{-1}Z'U) = 0,\; plim(p^{-1}Z'v) = 0 \tag{A.16}$$

Hence from [8, pp. 269 – 271], (A.11), (A.15) and (A.16):

$$plim(p^{-1}H) = G'VG \;\; where\; plim(p^{-1}U'U) \equiv V \equiv [\sigma_{kj}] \; for\; k, j \in J \tag{A.17}$$

$$= F'EF \;\; where\; E \equiv (\alpha', I)V(\alpha', I)' = [b_{1k}(\sigma_{11}b_{1j} + \sigma_{1j}) + \sigma_{k1}b_{1j} + \sigma_{kj}] \;\; for\; k, j \in J_1 \tag{A.18}$$

From (A.7), (A.18) and ([8], p. 271):

$$P^* \equiv plim(pP) = plim(p^{-1}H)^{-1} = F^{-1}D(F')^{-1} \;\; where\; D \equiv [d_{j\ell}] \equiv E^{-1} \tag{A.19}$$

From (A.7), (A.13), (A.16) and (A.19):

$$pR = -p(Z'Z)^{-1}(p^{-1}Z'U)GpP + CGpP,\; R^* \equiv plim(pR) = CGP^* = CGF^{-1}D(F')^{-1} \tag{A.20}$$

From (A.2), (A.4), (A.6), (A.11), (A.12) and (A.16):

$$\psi_1 = PX_0'v + QZ'v = P(G'U'v - G'C'Z'v) + QZ'v \; where\; \psi_1 \equiv [\hat{\beta}_k - \beta_k] \; for\; k = 2,...,n \tag{A.21}$$

$$\psi_2 = RX_0'v + SZ'v = R(G'U'v - G'C'Z'v) + SZ'v \;\; where\; \psi_2 \equiv [\hat{\beta}_{n+h} - \beta_{n+h}] \; for\; h = 1,...,m \tag{A.22}$$

$$plim(p^{-1}U'v) = \omega \;\; where\; \omega' \equiv -(\sigma_{11},...,\sigma_{1n}) \tag{A.23}$$

$$plim(p^{-1}X_0'v) = G'\omega = F'\omega_0 \;\; where\; \omega_0 \equiv [\alpha', I]\omega = -[b_{1\ell}\sigma_{11} + \sigma_{1\ell}]' \;\; for\; \ell \in J_1 \tag{A.24}$$

Hence from (A.5), (A.9) - (A.11), (A.16), (A.19) – (A.24):

$$\theta = (\theta^{o\prime}, \theta^{oo\prime})' \; for\; \theta^o \equiv plim\, \psi_1 = P^*G'\omega = F^{-1}D(F')^{-1}G'\omega = (B_0 + \varphi\alpha)D(F')^{-1}F'[\alpha', I]\omega \tag{A.25}$$

$$= (B_0 + \varphi\alpha)D\omega_0 = -[\sum_{j=2}^{n}(b_{kj} + \beta_k b_{1j})\sum_{\ell=2}^{n} d_{j\ell}(b_{1\ell}\sigma_{11} + \sigma_{1\ell})] \tag{A.26}$$

$$\theta^{oo} \equiv plim\ \boldsymbol{\psi_2} = \boldsymbol{R}^*\boldsymbol{G}'\boldsymbol{\omega} = \boldsymbol{CGF}^{-1}\boldsymbol{D}(\boldsymbol{F}')^{-1}\boldsymbol{F}'[\boldsymbol{\alpha}', \boldsymbol{I}]\boldsymbol{\omega} = \boldsymbol{C_0 D \omega_0}\ where\ \boldsymbol{C_0} \equiv \boldsymbol{C}[\boldsymbol{\alpha}', \boldsymbol{I}]' \tag{A.27}$$

$$= -[\sum_{j=2}^{n}(c_{hj} + c_{h1}b_{1j})\sum_{\ell=2}^{n} d_{j\ell}(b_{1\ell}\sigma_{11} + \sigma_{1\ell})] = -[\sum_{j=2}^{n}(b_{n+h,j} + \beta_{n+h}b_{1j})\sum_{\ell=2}^{n} d_{j\ell}(b_{1\ell}\sigma_{11} + \sigma_{1\ell})] \tag{A.28}$$

Proposition 1 follows directly from (A.25) - (A.28) and (A.18) - (A.19).

Proof of Proposition 2. Since the covariance matrix $\boldsymbol{V}$ associated with eqn (4) is symmetric and positive definite:

$$\boldsymbol{w}'\boldsymbol{V}\boldsymbol{w} > 0 \text{ for all } \boldsymbol{w} \neq \boldsymbol{0}, \text{ including } \boldsymbol{w} = (\boldsymbol{\alpha}', \boldsymbol{I})'\boldsymbol{y} \text{ for all } \boldsymbol{y} \neq \boldsymbol{0}; \text{ hence } \boldsymbol{y}'\boldsymbol{E}\boldsymbol{y} > 0 \text{ for all } \boldsymbol{y} \neq \boldsymbol{0} \tag{A.29}$$

using (A.18), so that the matrix $\boldsymbol{E} \equiv [e_{kj}]$ is also symmetric and positive definite. The matrix $-\boldsymbol{E}$ therefore has negative diagonal elements, and is a stable matrix (see [13, p. 165]), since $-\boldsymbol{E} - \boldsymbol{E}'$ is negative definite. By setting $\eta_{kk} = -e_{kk} + g$ for all $k \in J$ with $g = e_{k'k'} + \chi$, where $e_{k'k'}$ is the maximal diagonal element of $\boldsymbol{E}$ and χ is a positive constant, and by setting $\eta_{kj} = -e_{kj}$ for all off-diagonal elements, we may express $-\boldsymbol{E}$ in the form $-\boldsymbol{E} = \boldsymbol{N} - g\boldsymbol{I}$, where $\boldsymbol{N} \equiv [\eta_{kj}]$ is a Morishima matrix (see [1, p. 12]) under conditions (8) and (9). If $\boldsymbol{E}$ is also indecomposable, it follows from [13, pp. 213-215], that the elements $d_{j\ell}$ of $\boldsymbol{D} \equiv [d_{j\ell}] \equiv \boldsymbol{E}^{-1}$ are positive when $j, \ell \in J_1'$ and when $j, \ell \in J_1''$, and negative when $j \in J_1'\ \&\ \ell \in J_1''$ and when $j \in J_1''\ \&\ \ell \in J_1'$. Moreover when $\boldsymbol{E}$ has all non-zero elements and is not a 2x2 matrix, it follows from [12, p. 215] that conditions (8) and (9) are also necessary for the sign pattern of $\boldsymbol{D}$ to be determinate, given only the sign pattern of $\boldsymbol{E}$. Proposition 2 then follows from (7) in Proposition 1.

Proof of Proposition 3. From (A.3) and (A.4), we have:

$$\hat{\boldsymbol{v}}'\hat{\boldsymbol{v}} = \boldsymbol{v}'\boldsymbol{v} - \boldsymbol{v}'\boldsymbol{Y}_o(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\quad for\ \hat{\boldsymbol{v}} \equiv \boldsymbol{x} - \boldsymbol{Y_0}\hat{\boldsymbol{\beta}} = \boldsymbol{v} - \boldsymbol{Y_0}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \tag{A.30}$$

Hence using (A.2), (A.5), (A.16), (A.23) - (A.26):

$$plim\ s^2 = plim(1+(n/(p-n)))(plim(\boldsymbol{v}'\boldsymbol{v}/p) - plim(\boldsymbol{v}'\boldsymbol{Y}_o/p)plim(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}))\ for\ s^2 \equiv \hat{\boldsymbol{v}}'\hat{\boldsymbol{v}}/(p-n) \tag{A.31}$$

$$= \sigma_1^2 - \boldsymbol{\omega_0}'\boldsymbol{F}\boldsymbol{\theta}^o = \sigma_1^2 - \boldsymbol{\omega}_0'\boldsymbol{D}\boldsymbol{\omega}_o\ \ where\ \ \boldsymbol{\omega}_0'\boldsymbol{D}\boldsymbol{\omega}_o > 0\ for\ \boldsymbol{\omega}_o \neq 0 \tag{A.32}$$

with $\boldsymbol{E}$ in (A.29), and hence $\boldsymbol{D}$ in (A.19), positive definite. Eqn (16) in Proposition 3 follows from (A.32) and (A.24), and eqn (17) from (A.32) and (A.33) - (A.34) below when $\boldsymbol{V}$ is diagonal.

Proof of Proposition 4. We may show using (A.18) that if $\boldsymbol{V} = [\delta_{kj}\sigma_{jj}]$ for all $k, j \in J$, where δ_{kj} is Kronecker's delta, so that $\boldsymbol{V}$ is a diagonal matrix:

$$\boldsymbol{E} = (\boldsymbol{\alpha}'\sigma_j^2\boldsymbol{\alpha} + \boldsymbol{V_0}) \text{ and } \boldsymbol{D} \equiv \boldsymbol{E}^{-1} = [(\delta_{kj}/\sigma_k^2) - ((b_{1j}b_{1k}\sigma_1^2/\sigma_j^2\sigma_k^2\zeta))] \tag{A.33}$$

$$where\ \boldsymbol{V}_0 \equiv [\delta_{kj}\sigma_{jj}]\ for\ k, j\varepsilon J_1 \quad and\ \zeta \equiv (1+\sum_{j=2}^{n} b_{1j}^2(\sigma_1^2/\sigma_j^2))>0 \tag{A.34}$$

with $\boldsymbol{D}(\alpha'\sigma_j^2\,\alpha + \boldsymbol{V_0}) = \boldsymbol{I}$, given (A.9). From (A.23)-(A.24), (A.33) and (A.34), if $\boldsymbol{V}$ is diagonal:

$$\boldsymbol{\omega} \equiv -(\sigma_1^2, 0, ..,0)',\ \boldsymbol{\omega_0} = -[b_{1\ell}\sigma_1^2]'\ \ for\ \ell\varepsilon J_1,\ \ \boldsymbol{D\omega_0} = -[b_{1j}\sigma_1^2/\zeta\sigma_j^2]'\ \ for\ j\varepsilon J_1 \tag{A.35}$$

Hence from (A.5), (A.25) – (A.28) and (A.36):

$$\theta_k \equiv plim\ \hat{\beta}_k - \beta_k = -\frac{\sigma_1^2}{\zeta}\sum_{j\in J_1} b_{1j}(\beta_k b_{1j} + b_{kj})/\sigma_j^2 \quad for\ k = 2,...,n+m \tag{A.36}$$

Proof of Propositions 5 – 10. These follow directly from (A.34), (A.36) and the stability condition (20).

Proof of Proposition 11. The OLS t-statistic for testing the null hypothesis that $\beta_k = 0$ is given (see e.g. [8, p. 182]) by:

$$t_k = \hat{\beta}_k/(s\varsigma_{kk}^{0.5})\ with\ t_k' \equiv plim\ t_k = \hat{\beta}_k^o/(s_o\tilde{\varsigma}_{kk}^{0.5})\ \ for\ s_o \equiv plim\ s \tag{A.37}$$

for any given set of observations $\boldsymbol{Y_0}$ and the associated diagonal elements ς_{kk} of $(\boldsymbol{Y_0'Y}_0)^{-1}$, with $\tilde{\varsigma}_{kk} \equiv plim\ \varsigma_{kk}$. If the OLS estimates were unbiased, we would have:

$$\hat{\beta}_k^o = \beta_k\ \ and\ s_o = \sigma_1\ with\ plim\ t_k = t_{ok} \equiv \beta_k/(\sigma_1\tilde{\varsigma}_{kk}^{0.5}) \tag{A.38}$$

However, eqns (16), (17), (27) and (29) imply that for $\beta_k \neq 0$, when $J_0 \neq \Phi$:

$$plim\ t_k = t_k' = \hat{\beta}_k^o\zeta^{0.5}/(\sigma_1\,\tilde{\varsigma}_{kk}^{0.5})\ with\ (t_k'/t_{ok}) = \hat{\beta}_k^o\zeta^{0.5}/\beta_k = (1-\sigma_1^2\sum_{j\in J_0}(b_{1j}b_{kj}/\beta_k\sigma_j^2))/\zeta^{0.5} \tag{A.39}$$

implying that the asymptotic proportionate bias, $(t_k' - t_{ok})/t_{ok}$ is given by (26) in Proposition 11.

Proof of Propositions 12 – 13. These follow directly from eqns (16), (17), (27) and (29).

Proof of Propositions 14 – 17. These follow from eqn (16) and differentiation of eqn (17).

Proof of Proposition 18. If we define θ_k^r as the value of the bias $plim\ \hat{\beta}_k - \beta_k$ when equation $r \in J_1$ is not included in the set of simultaneous equations in (4), with $J_{1r} \equiv J_1 - \{r\}$, we have from (A.36), (27) and (29):

$$\Delta_k \equiv \theta_k - \theta_k^r = (\theta_k(\zeta_r - \zeta) - (\sigma_1^2 b_{1r}(\beta_k b_{1r} + b_{kr})/\sigma_r^2))/\zeta_r \tag{A.40}$$

$$= -\sigma_1^2 b_{1r}(b_{1r}\hat{\beta}_k^0 + b_{kr})/(\zeta_r \sigma_r^2) \;\; where\; \zeta_r \equiv (1 + \sum_{j \in J_{1r}} b_{1j}^2(\sigma_1^2/\sigma_j^2)) > 1 \qquad \text{(A.41)}$$

Proposition 18 follows directly from (A.40) – (A.41)

Proof of Proposition 19. Let $\boldsymbol{Z_1}$ be the matrix composed of the first m' columns of $\boldsymbol{Z}$, $\boldsymbol{Z_2}$ the matrix composed of the last $m-m'$ columns of $\boldsymbol{Z}$, $\boldsymbol{Y_1} \equiv [\boldsymbol{X}, \boldsymbol{Z_1}]$ and $\hat{\boldsymbol{\beta}}''$ the OLS estimate of $\boldsymbol{\beta}'' \equiv [\beta_k]$ for $\mathrm{k} = 2,..., \mathrm{n}+\mathrm{m}'$, with $\boldsymbol{\beta}''' \equiv [\beta_k]$ for $\mathrm{k} = \mathrm{n}+\mathrm{m}'+1,...,\mathrm{n}+\mathrm{m}$. We then have:

$$plim\; \hat{\boldsymbol{\beta}}'' = plim(\boldsymbol{Y_1'Y_1})^{-1}\boldsymbol{Y_1'}(\boldsymbol{Y_1\beta''} + \boldsymbol{Z_2\beta'''} + \boldsymbol{v}) = \boldsymbol{\beta''} + \boldsymbol{T^*\beta'''} + plim(\boldsymbol{Y_1'Y_1})^{-1}\boldsymbol{Y_1'v} \qquad \text{(A.42)}$$

for $\boldsymbol{T^*} \equiv plim(\boldsymbol{Y_1'Y_1})^{-1}\boldsymbol{Y_1'Z_2}$. Since from (A.2), $\boldsymbol{Y_0} \equiv [\boldsymbol{Y_1}, \boldsymbol{Z_2}]$:

$$\boldsymbol{Y_0'Y_0} = \begin{pmatrix} \boldsymbol{Y_1'Y_1} & \boldsymbol{Y_1'Z_2} \\ \boldsymbol{Z_2'Y_1} & \boldsymbol{Z_2'Z_2} \end{pmatrix} \text{with } (\boldsymbol{Y_0'Y_0})^{-1} \equiv \begin{pmatrix} \boldsymbol{P_1} & \boldsymbol{Q_1} \\ \boldsymbol{R_1} & \boldsymbol{S_1} \end{pmatrix} \text{ and hence } \boldsymbol{Y_1'Y_1P_1} + \boldsymbol{Y_1'Z_2R_1} = \boldsymbol{I} \qquad \text{(A.43)}$$

where $\boldsymbol{P_1}$ is $(n+m')$x$(n+m')$, $\boldsymbol{Q_1}$ is $(n+m')$x$(m-m')$, $\boldsymbol{R_1}$ is $(m-m')$x$(n+m')$, and $\boldsymbol{S_1}$ is $(m-m')$x$(m-m')$. Since the predetermined variables in $\boldsymbol{Z_2}$ are uncorrelated with $\boldsymbol{v}$, $plim\boldsymbol{Z_2'v}$. $= \boldsymbol{0}$. Hence from (A.42) and (A.43):

$$plim(\boldsymbol{Y_1'Y_1})^{-1}\boldsymbol{Y_1'v} = \boldsymbol{\theta^o} + \boldsymbol{T^*\theta^{oo}} \; where\; \boldsymbol{\theta^o} \equiv plim(\boldsymbol{P_1Y_1'v} + \boldsymbol{Q_1Z_2'v}), \boldsymbol{\theta^{oo}} \equiv plim(\boldsymbol{R_1Y_1'v} + \boldsymbol{S_1Z_2'v}) \qquad \text{(A.44)}$$

From (A.5) and (A.43), $\boldsymbol{\theta^o}$ corresponds to the first $n+m'-1$ rows, and $\boldsymbol{\theta^{oo}}$ to the last $m-m'$ rows, of the asymptotic bias $\boldsymbol{\theta}$ in (A.5) that results when all $n+m$ variables are included in the OLS regression. (A.42) and (A.44) in turn imply equation (35) in Proposition 19.

We may re-write equation (4) in the form:

$$\boldsymbol{XB} + \boldsymbol{Z_1C_1} = \boldsymbol{U_o} \;\; where\;\; \boldsymbol{U_o} \equiv \boldsymbol{U} - \boldsymbol{Z_2C_o} \qquad \text{(A.45)}$$

$\boldsymbol{C_1}$ consists here of the first m' rows, and $\boldsymbol{C_o}$ of the last $m-m'$ rows, of $\boldsymbol{C}$. The covariance matrix associated with $\boldsymbol{U_o}$ is $\boldsymbol{V_o} \equiv \boldsymbol{V} + \boldsymbol{C_o'\Omega_oC_o}$, where $\boldsymbol{\Omega_o}$ is the $(m-m')$ x $(m-m')$ (positive definite) covariance matrix for the $m-m'$ excluded predetermined variables. (A.45) is now of the same form as (4), but with a new covariance matrix $\boldsymbol{V_o}$ in place of the original covariance matrix $\boldsymbol{V}$ for the RHS disturbance terms, and with the last $m-m'$ predetermined variables excluded from the LHS of the first equation and all other equations. Eqns (36) and (37) of Proposition 19 then follow from a parallel application of Propositions 1 and 4 respectively to this transformed equation system.

Proof of Proposition 20. Using (38), we may express eqn (4) in the form:

$$\boldsymbol{X_1B_{11}} + \boldsymbol{X_2B_{21}} + \boldsymbol{Z_1C_{11}} + \boldsymbol{Z_2C_{21}} = \boldsymbol{U_1},\;\; \boldsymbol{X_1B_{12}} + \boldsymbol{X_2B_{22}} + \boldsymbol{Z_1C_{12}} + \boldsymbol{Z_2C_{22}} = \boldsymbol{U_2} \qquad \text{(A.46)}$$

where $\boldsymbol{U_1}$ contains the first n' rows of $\boldsymbol{U}$ and $\boldsymbol{U_2}$ the remaining $n-n'$ rows of $\boldsymbol{U}$, and $\boldsymbol{X_1}$ contains the first n', and $\boldsymbol{X_2}$ the remaining, $n-n'$ columns of $\boldsymbol{X}$. Hence:

$$\boldsymbol{X_2} = -\boldsymbol{X_1 B_{12} B_{22}^{-1}} - \boldsymbol{Z_1 C_{12} B_{22}^{-1}} - \boldsymbol{Z_2 C_{22} B_{22}^{-1}} + \boldsymbol{U_2 B_{22}^{-1}} \qquad \text{(A.47)}$$

if $\boldsymbol{B_{22}}$ is non-singular. Substituting (A.47) into the first part of (A.46) yields:

$$\boldsymbol{X_1 B''_{11}} + \boldsymbol{Z_1 C''_{11}} = \boldsymbol{U''_1} \ \ \textit{where} \ \ \boldsymbol{U''_1} \equiv (\boldsymbol{U_1} - \boldsymbol{Z_2}(\boldsymbol{C_{21}} - \boldsymbol{C_{22} A_1}) - \boldsymbol{U_2 A_1})\boldsymbol{A_0} \ \ \textit{for} \ \boldsymbol{A_1} \equiv \boldsymbol{B_{22}^{-1} B_{21}} \qquad \text{(A.48)}$$

$$\boldsymbol{B''_{11}} \equiv [b''_{k,j}] \equiv (\boldsymbol{B_{11}} - \boldsymbol{B_{12} A_1})\boldsymbol{A_o} \text{ for } k, j \in J', \boldsymbol{C''_{11}} \equiv (\boldsymbol{C_{11}} - \boldsymbol{C_{12} A_1})\boldsymbol{A_o} \qquad \text{(A.49)}$$

and post-multiplication by $\boldsymbol{A_o} \equiv [\delta_{kj} / \vartheta_j]$ for $k, j \in J'$, where ϑ_j is the element in the jth row and jth column of the matrix $(\boldsymbol{B_{12} A_1} - \boldsymbol{B_{11}})$, ensures that $b''_{jj} = -1$ for all $j = 2,...,n'$. (A.48) is now of the same form as (4), but with an associated covariance matrix for $\boldsymbol{U''_1}$ given by $\boldsymbol{V''}$ in (39). We can now apply parallel versions of Proposition 1, and of Proposition 4 if $\boldsymbol{V''}$ is diagonal, to (A.48), yielding Proposition 20.

References

[1] M. Allingham, M. Morishima, Qualitative economics and comparative statics, in: M. Morishima et al, Theory of Demand, Oxford UP, Oxford, 1973, pp. 3-69.

[2] R. Davidson, J. MacKinnon, Econometric Theory and Methods, Oxford UP, New York, 2004.

[3] G. Hadley, Linear Algebra. London: Addison Wesley, London, 1961.

[4] E. Hanushek, The economics of schooling: production and efficiency in public schools, J. Econ. Lit. 24 (1986) 1141 – 1177.

[5] E. Hanushek, Assessing the effects of school resources in student performance: an update, Educ. Evaluation and Policy Anal. 19 (1997) 141 – 164.

[6] J. Hausman, Specification tests in econometrics, Econometrica, 46 (1978) 1251-1271.

[7] J. R. Hicks, Value and Capital, 2nd edn. , Oxford University Press, Oxford, 1946.

[8] J. Johnston, Econometric Methods, 3rd edn., McGraw-Hill, New York, 1984.

[9] J. R. Magnus, A. Vasnev, Local sensitivity and diagnostic tests, Econometrics J., 10 (2007) 166-192.

[10] C. Mooney, Monte Carlo Simulation, Sage Publications, London, 1997.

[11] A. Nakamura, M. Nakamura, On the performance of tests by Wu and by Hausman for detecting the ordinary least squares bias problem, J. Econometrics, 29 (1985) 213-227.

[12] A. Nakamura, M. Nakamura, Model specification and endogeneity, J. Econometrics, 83 (1998) 213-237.

[13] J. Quirk, R. Saposnik, Introduction to General Equilibrium Theory and Welfare Economics, McGraw-Hill, New York, 1968.

[14] D. Wu, Alternative tests of independence between stochastic regressors and disturbances, Econometrica, 41 (1973) 733-750.