SENSITIVITY ANALYSIS OF OLS MULTIPLE REGRESSION INFERENCE WITH RESPECT TO POSSIBLE LINEAR ENDOGENEITY IN THE EXPLANATORY VARIABLES

RICHARD A. ASHLEY AND CHRISTOPHER F. PARMETER

Abstract. This work describes a versatile sensitivity analysis of OLS hypothesis test rejection p-values with respect to possible endogeneity in the explanatory variables of the usual k-variate linear multiple regression model which practitioners can readily deploy in their research. This sensitivity analysis is based on a derivation of the asymptotic distribution of the OLS parameter estimator, but extended in a particularly straightforward way to the case where some or all of the explanatory variables are endogenous to a specified degree – that is, where the population covariances of the explanatory variables with the model errors are given. In exchange for restricting attention to possible endogeneity which is solely linear in nature, no additional model assumptions must be made, beyond the usual ones for a model with stochastic regressors. In addition, we also use simulation methods to quantify the uncertainty in the sensitivity analysis results introduced by replacing the population variance-covariance matrix by its sample estimate. The usefulness of the analysis – as a 'screen' for potential endogeneity issues – is illustrated with an example from the empirical growth literature.

JEL Classification: C2, C15.

1. Introduction

For many of us, the essential distinction between econometric regression analysis and otherwise-similar forms of regression analysis conducted outside of economics, is the overarching concern shown in econometrics with regard to the model assumptions made in order

Date: January 16, 2019.
Key words and phrases. Robustness, Exogeneity, Multiple Regression, Inference, Instrumental Variables.
Richard Ashley, Corresponding Author, Department of Economics, Virginia Polytechnic Institute and State University; e-mail: ashleyr@vt.edu. Christopher Parmeter, Department of Economics, University of Miami; e-mail: cparmeter@bus.miami.edu. We are indebted to Professor Kiviet for pointing out (Kiviet, 2016) several deficiencies in Ashley and Parmeter (2015a). In particular, we erred in failing to acknowledge the previous work of Kiviet and Niemczyk (2007, 2012) and Kiviet (2013) on the sampling distribution of the OLS estimator in a regression with an endogenous explanatory variable. We also erred in failing to address the impact of sampling error in estimating the variances of the regression explanatory variates on the asymptotic distribution of the OLS coefficient estimator. In response to his critique, we here provide a substantively fresh approach to this topic. Moreover, because of his effort, we were able to not only correct our distributional result, but also obtain a deeper and more intuitive analysis; the new analytic results obtained here are an unanticipated bonus.

to obtain consistent OLS parameter estimation and asymptotically valid statistical inference results. As Friedman (1953) famously noted (in the context of economic theorizing), it is both necessary and appropriate to make model assumptions – notably, even assumptions which we know to be false – in any successful economic modeling effort: the usefulness of a model, he asserted, inheres in the richness/quality of its predictions rather than in the accuracy of its assumptions. Our contribution here – and in Ashley and Parmeter (2015b), which addresses similar issues in the context of GMM/IV inference using possibly-flawed instruments – is to both posit and operationalize a general proposition that is a natural corollary to Friedman's assertion: It is perfectly all right to make possibly-false (and even very-likely-false) assumptions – *if and only if one can and does show that the model results one most cares about are insensitive to the levels of violations in these assumptions it is reasonable to expect*. The present context – in which we address the possibility of endogeneity in one's multivariate linear regression model explanatory variables – provides an ideal setting in which to both exhibit and operationalize a quantification of the 'insensitivity' alluded to in this proposition, because this setting is so very simple. This setting is also attractive in that OLS estimation of multiple regression models with explanatory variables of suspect exogeneity is very common in applied economic work.

The present paper proposes a sensitivity analysis for OLS estimation/inference in the presence of unmodeled endogeneity in the explanatory variables of multiple regression models. Earlier work on this subject (Kiviet and Niemczyk, 2007, 2012; and Kiviet, 2013, 2016) has stimulated us to obtain essentially analytic results for the important special case where the sensitivity with respect to the possible endogeneity of a single explanatory variable is being examined. Beyond this special case, our results here represent a general, and easily deployed, approach for assessing how credible one's inferential findings are to (solely) linear dependence between the covariates of the regression model and the unobserved model error term. The implications of this linear dependence restriction are discussed in Section 2 below.

The sensitivity analysis proposed here is designed to be so straightforward an addendum to the OLS estimation applied economists are already doing – both theoretically and in terms of the effort required to set up and use the procedure – that analysts will actually, routinely, use it. In this regard we see our sensitivity analysis as what could be characterized as a 'screen' for possibly-important unaddressed endogeneity issues.

Thus, in particular, the theoretical development underlying our sensitivity analysis (which is mostly embodied in our derivation below of the sampling distribution of the OLS parameter estimators for the usual k-variate multiple regression model, when some or all of the explanatory variables are endogenous to a specified degree) is very substantially simplified by the restriction we make here to what we are terming 'linear endogeneity.' Linear endogeneity simply means that any relationship between the explanatory variables in the regression model and the model error term is solely linear, and hence completely captured (quantified) by the relevant covariances or correlations between the explanatory variables and the model errors. This kind of endogeneity is clearly the sort which analysts generally have in mind, since endogeneity is ordinarily described in terms of correlation between explanatory variates and the model errors. And it is a natural restriction to place on the scope of a sensitivity analysis in the present (linear regression) context, since practically any form of endogeneity can be viewed as the existence of a set of explanatory variables which have been wrongly-omitted from the model specification, but which are correlated with one or more of the included explanatory variables. Thus, in this context, 'linear endogeneity' merely specifies that these wrongly-omitted variables are linear functions of the already-included explanatory variables. Clearly, this is a restriction, however: some or all of the wrongly-omitted variables could in principle enter the fully-correct model specification in a nonlinear fashion. Still, one might expect that linear endogeneity would subsume most of any endogeneity actually present, so that a sensitivity analysis restricted to quantifying the impact of possible linear endogeneity should be a good approximation to a sensitivity analysis not restricted in this way. And – as will be apparent in the sampling distribution derivation given in Section 2 – this restriction

turns a challenging sampling distribution derivation into one that is so brief and so completely straightforward as to be easily accessible to any potential user of the technique. We return to this topic immediately below, and again in Section 2.

In addition we have designed our procedure to be very easy for a user to start up with and utilize, making the 'trouble' involved with adopting it as a routine screen for quantifying the impact of possible endogeneity problems quite a small barrier to its adoption and use. In particular, our sensitivity analysis procedure is already implemented in `R` and `Stata` scripts, and hence does not ask the user to invest in learning/implementing a new estimation framework. And – importantly – it requires no additional model assumptions on the regression model at issue, beyond the ones which would ordinarily need to be made in OLS multiple regression analysis with stochastic regressors. In particular, our approach does not require the user to make any additional assumptions about the distribution of the model errors: one need not assume that these are Gaussian, nor need one make any (empirically inaccessible) assumptions as to the values of any of their higher moments. Our approach does restrict attention to possible endogeneity which is only linear in form, but the meaning of this assumption is so clearly understandable that the user is in a position to judge its restrictiveness for himself/herself. For our part, we note that it hardly seems likely in practice that endogeneity at a level that is of practical importance will arise which does not engender a notable degree of correlation (i.e., linear relationship) between the explanatory variables and the model error term.

Returning to the issue of practical usability, all that the user needs to provide is (a) a list of which explanatory variables are under consideration in the sensitivity analysis as being possibly endogenous and (b) a choice of a particular null hypothesis whose robustness or fragility with respect to this possible endogeneity is of interest. Our implementing software then calculates what we call 'r_{min}', where r_{min} is defined to be the minimum-length vector of

correlations – between the set of explanatory variables under consideration and the (unobserved) model errors – which suffices to 'overturn' the observed rejection (or non-rejection) of the chosen null hypothesis.[1]

The length of this correlation vector (i.e., $|r_{min}|$) quantifies the robustness of this particular inference to possible linear endogeneity in these explanatory variables: if it is large (e.g., greater than, say 0.40 or 0.50) then a reasonable person would surely designate this inference result as 'robust' with respect to possible endogeneity in this set of explanatory variables; whereas a small value for this length (such as $|r_{min}| < 0.05$) would certainly imply that this particular inference result is 'fragile' with respect to possible endogeneity in these explanatory variables. True, the robustness/fragility conclusion one can reach on the basis of this calculated $|r_{min}|$ value is clearly subjective for intermediate values of this r_{min} length. But $|r_{min}|$ does provide objective sample evidence for addressing this conclusion; this evidence is objective in the sense that any analyst using these data would obtain the same value for $|r_{min}|$.

This evidentiary situation (with regard to the robustness or fragility conclusion one might base on $|r_{min}|$) is analogous to the objective (and subjective) aspects of the situation which obtains with respect to rejection of an ordinary null hypothesis on the basis of an observed rejection p-value: This p-value is objective, in that any analyst using the same model and data will obtain the same value for it; yet the conclusion based on a p-value is subjective, in that the actual decision as to whether or not rejection of the null hypothesis is warranted, based on this evidence against it, is inherently and appropriately a 'judgement call.'

[1]Flexible implementing software in both `R` and `Stata` is available from the authors; these scripts can be used as standalone programs or simply patched into the `R` or `Stata` code the analyst is already using for estimating their model(s). Given the data (e.g., input in csv format), a list of the explanatory variables which are considered to be possibly endogenous, and the specification of the null hypothesis which was rejected at, say, the 5 percent significance level (or which failed to be rejected at this level), the software calculates r_{min} for this null hypothesis. Our routines can handle any sort of null hypothesis for which `R` or `Stata` can compute a rejection p-value (simple or compound, linear or nonlinear) using whatever kind of standard error estimates (OLS/White-Eicker/HAC) are already being used. They can also, optionally, implement the bootstrap simulations with respect to the explanatory variable variances described below.

If the sensitivity analysis 'screen' proposed here shows that the inferences one most cares about are robust with respect to reasonable amounts of endogeneity, then one can dispense with a search for valid instruments. In contrast, where the sensitivity analysis indicates that these key inference results are fragile in that regard, then the motivation for such an instrument search might become quite strong. Such a search – if feasible at all – almost invariably leads to instruments of at least somewhat questionable validity – that is, the exogeneity of these proposed instruments is itself in at least some doubt. Thus, in using IV inference the issue naturally arises as to whether the resulting instrumental-variables versions of the analyst's key inferences are themselves potentially (similarly) fragile – or robust – with respect, now, to possible instrument invalidity. (A sensitivity analysis for that context, analogous to that proposed here, was introduced and operationalized in Ashley and Parmeter, 2015b). Alternatively, in this case where the sensitivity analysis 'screen' proposed here shows that the inferences one most cares about are fragile with respect to reasonable amounts of endogeneity (and where IV inference is itself either infeasible or fragile), then the analyst might reasonably be motivated to make the additional assumptions and/or investments needed in order to go beyond OLS estimation and use one of the estimation procedures proposed in Caner and Morrill (2013), Kraay (2012), or Kiviet (2018); Or perhaps – if there is strong evidence for heteroscedasticity – one might use the instruments suggested in Lewbel (2012).

In Section 2 below we derive the asymptotic sampling distribution of the OLS k-variate multiple regression model parameter estimator, under the usual assumptions for such models with stochastic regressors except explicitly allowing the k explanatory variates to be linearly endogenous, each with a specified covariance with the model error. As a check on our results, we then show that this sampling distribution reduces to the Kiviet (2016) result for the special case of a bivariate regression model with a single (endogenous) explanatory variable.

In Section 3 we utilize this sampling distribution in a sensitivity analysis analogous to that proposed in Ashley and Parmeter (2015a,b) for IV estimation/inference with possibly-flawed instruments, showing how straightforward the required calculations are in general and deriving analytic results for the special - but not uncommon - case where only a single explanatory variable is considered to be possibly-endogenous and the null hypothesis being tested is a single linear restriction on the structural coefficients.

Section 4 applies these new results to the sensitivity analysis results on several inferences in the Mankiw, Romer and Weil (1992) estimated aggregate production function. These results provide a nice illustration for the present paper in that one of MRW's key null hypotheses is a simple zero-restriction and the other is a linear restriction on their model coefficients; also we find that some of their inferences are robust with respect to possible endogeneity in their explanatory variables, whereas others are fragile. Importantly, we use bootstrap simulation to provide an estimated standard error for each $|r_{min}|$ value estimated. These simulations explicitly allow for the impact on the asymptotic sampling distribution of the OLS structural coefficient estimator which arises from the substitution of sample estimates of the variances of the explanatory variables into the derived asymptotic bias expression, which depends on the analogous population variances of these variables. This MRW example illustrates the usefulness of the sensitivity analysis proposed here.

2. The OLS Multiple Regression Estimator With Endogenous Covariates

2.1. **Sampling Distribution of the OLS Estimator.** Assume that

$$Y = X\beta + \varepsilon, \tag{1}$$

where the matrix of regressors, X, is $n \times k$ with (for simplicity) zero mean and population variance-covariance Σ_{XX}. Here $\ell \leq k$ of the explanatory variables are taken to be 'linearly endogenous' – i.e., related to the error term ε, but in a solely linear fashion – with covariance $E(\frac{1}{n}X'\varepsilon)$ given by the k-vector λ. Then ε can be expressed as a linear function of X plus a

random variate ν which is – by construction – independent of X:

$$\varepsilon = X\Sigma_{XX}^{-1}\lambda + \nu, \tag{2}$$

where ν has mean zero and variance σ_ν^2.[2]

Substituting Equation (2) into Equation (1) results in

$$Y = X\beta + \varepsilon = X\beta + (X\Sigma_{XX}^{-1}\lambda + \nu) = X(\beta + \Sigma_{XX}^{-1}\lambda) + \nu. \tag{3}$$

This is the regression model one is actually estimating, and – under our restriction to consider solely-linear endogeneity – satisfies all of the usual assumptions necessary for consistent OLS parameter estimation and asymptotically valid inference with stochastic regressors, as these regressors are now forced to be independent of the new model error term, ν.

Were X and ν in addition iid, then the sampling distribution of $\widehat{\beta}^{OLS}$ is thus (e.g., from Johnston, 1972 Section 9-2):

$$\sqrt{n}\left(\widehat{\beta}^{OLS} - (\beta + \Sigma_{XX}^{-1}\lambda)\right) \xrightarrow{D} N(0, \sigma_\nu^2\Sigma_{XX}^{-1}). \tag{4}$$

The value of σ_ν^2 can in that case be consistently estimated from the OLS fitting errors – using the usual estimator, s^2 – but Equation (4) makes it plain that $\widehat{\beta}^{OLS}$ is biased and inconsistent if and only if $\lambda \neq 0$.[3] The asymptotic variance of $\widehat{\beta}^{OLS}$ given by Equation (4) would no longer remain under less restrictive assumptions on (X, ν) but one would in that case use White-Eicker or Newey-West standard errors.

The restriction here to the special case of solely-linear endogeneity thus makes the derivation of at least the asymptotic mean of the sampling distribution of $\widehat{\beta}^{OLS}$ almost trivial. But

[2]Multiplication of (2) by $n^{-1}X'$ and taking expectations yields $E(n^{-1}X'\varepsilon) = E\left[n^{-1}X'X\Sigma_{XX}^{-1}\lambda + n^{-1}X'\nu\right] = \lambda + E(n^{-1}X'\nu) = \lambda + 0$, which verifies the form of (2).

[3]Also, Σ_{XX}^{-1} can be consistently estimated from the sample data on the explanatory variables. The reader is cautioned, however, that – as pointed out in Kiviet (2016) – sampling variation in the sample estimate of Σ_{XX} has a impact on the asymptotic distribution of the resulting $\widehat{\beta}^{OLS}$ estimator, through its influence on the estimated asymptotic bias term. This complication is explicitly addressed in the sensitivity analysis algorithm detailed in Section 3 below, using bootstrap simulation.

it is legitimate to wonder what this restriction is approximating away, relative to instead – as in Kiviet (2016), say – making an *ad hoc* assumption about the values of the fourth moments of the variables (X, ε). It is well known that Gaussian variates are either linearly related or not related at all, so an assumption of joint Gaussanity for (X, ε) would clearly suffice to imply that any endogeneity in this explanatory variable is solely linear, and ν would be independent of X in that case. It is worth noting, however, that many other distributional assumptions on (X, ε) still lead – as in Equation (3) – to $E(Y|X) = X\beta$, so that X and ν are at least uncorrelated in those cases. For example, where (X, ε) is distributed jointly as Student's t, Spanos (1999, pp. 344 and 373-374) has shown that $E(Y|X)$ is, as in the joint Gaussian case, a linear function of X. So this distributional assumption, too, implies that Equation (4) is a valid regression equation, albeit (in this case) with an error term exhibiting heteroscedasticity which is quadratic in X. Presumably many other (un-named and un-tabulated) joint distributions for (X, ε) similarly imply that $E(Y|X)$ is a linear function of X; in those cases the heteroscedasticity in ν is of unknown form, but this is immaterial once one is using White-Eicker standard error estimates.

Our restriction here to a consideration only of possible endogeneity which is linear in form is in this sense substantially less restrictive than the particular fourth moment assumption corresponding to joint Gaussianity. On the other hand, it clearly *is* a restriction: simply assuming that ν (the error term left over once the linear dependence of ε on X has been stripped out) is independent of X could be an assumption which is non-trivially incorrect, leading to inference (and sensitivity analysis) results which are to some degree distorted. Such distortions could be avoided by using asymptotic results based on the correct values of all of the (X, ε) fourth moments. But these fourth moments are in practice quite difficult to estimate with any accuracy; and those involving ε are outright infeasible to estimate, since ε is not observable. Any results one would obtain based on inaccurately estimated (or *ad hoc*) fourth moment estimates would of course be at least somewhat distorted also. At least the assumption of linear endogeneity has a clear economic meaning: in this linear regression

analysis we are assuming that the omitted explanatory variables that are in any way related to the included ones are related to these included variates in an essentially linear manner, which seems appropriate to a linear regression modeling enterprise.

Further, it follows from Equation (2) that

$$\begin{aligned} \varepsilon'\varepsilon &= \left(X\Sigma_{XX}^{-1}\lambda + \nu\right)' \left(X\Sigma_{XX}^{-1}\lambda + \nu\right) \\ &= \lambda'\Sigma_{XX}^{-1}X'X\Sigma_{XX}^{-1}\lambda + 2\lambda'\Sigma_{XX}^{-1}X'\nu + \nu'\nu. \end{aligned} \tag{5}$$

Dividing both sides of this equation by n and taking expectations yields

$$\sigma_\varepsilon^2 = \lambda'\Sigma_{XX}^{-1}\lambda + \sigma_\nu^2 > \sigma_\nu^2. \tag{6}$$

Consequently, an additional useful feature of the 'linear endogeneity' framework used here is that the simple derivation above transparently exhibits the essence of the impact of explanatory variable endogeneity on OLS estimation and inference: Where explanatory variables in the original ('structural') Equation (1) are endogenous (e.g., because of wrongly-omitted variates which are correlated with the included ones), then this endogeneity actually improves the fit of the estimated model, Equation (3). This takes place simply because Equation (3) is using some or all of its included variables to also model ('proxy for') the 'omitted variables' portion of the structural error, ε; this proxying can take place if and to the extent that the omitted variables are correlated with these included variables, which is of course the source of the endogeneity in these included variables. Thus, the actual OLS fitting error (ν) has a smaller variance than does the Equation (1) structural error (ε), but the resulting OLS parameter estimator is then inconsistent for the Equation (1) structural parameter, β.

This proxying is actually beneficial in a forecasting context, where explanatory variables are being wrongly omitted because the sample data for them are unavailable: in that setting it is a good thing for the estimated forecasting equation to embody both the direct impacts of the observed explanatory variables and also their indirect forecasting power via their

correlations with the unavailable forecasting variables. But this inconsistent β estimation is clearly deleterious to the testing of null hypotheses involving the structural parameter vector, β itself, as defined by Equation (1).

Up to this point λ, the vector of covariances of the k explanatory variables with the structural model errors, ε, has been taken as given. However – absent a consistent estimator of β – these structural errors are inherently unobservable, even asymptotically, unless one has additional (exogenous) information not ordinarily available in empirical settings. Consequently λ is inherently not estimable . It *is*, however, possible – quite easy, actually – to estimate how sensitive the rejection p-value for any particular null hypothesis involving β is with respect to correlations between the explanatory variables and these unobservable structural errors. The next section of this paper describes our algorithm for accomplishing this sensitivity analysis, which is the central result of the present paper.

Finally, it useful to note at the outset that this sensitivity analysis algorithm – based on the simple $\widehat{\beta}^{OLS}$ sampling distribution derived above as Equation (4) – is in no way daunting to actually implement, as we have embodied it into both `R` and `Stata` scripts. These scripts can be easily inserted into the estimation code one is already using, and readily applied with respect to whatever null hypothesis one is already testing, whether it be a single linear restriction or a set of multiple linear restrictions. In practical applications the fitting errors for Equation (3) – the regression model actually estimated – might well show indications of heteroscedasticity and/or serial correlation. While model re-specification is the best response to such indications, where this is either not possible or where such re-specification fails to produce the iid model errors assumed above, our scripts allow one to replace $s_\nu^2\Sigma_{XX}^{-1}$ – the sample estimate of $\sigma_\nu^2\Sigma_{XX}^{-1}$ in Equation (4) – by the corresponding (consistent) White-Eicker or Newey-West estimates.

2.2. **Relation to the Single-Regressor Result of Kiviet (2016).** For the bivariate regression special case explicitly considered by Kiviet (2013, 2016), the $k \times k$ variance-covariance matrix Σ_{XX} reduces to the scalar σ_x^2 and the k-vector λ reduces to the scalar

$$\lambda = Cov(x_i, \varepsilon_i) = \sigma_{x\varepsilon} = \rho_{x\varepsilon}\sqrt{\sigma_x^2 \sigma_\varepsilon^2}, \tag{7}$$

for all $i \in [1, \ldots, n]$, where x_i and ε_i each denote the vector's i^{th} component, and where $\rho_{x\varepsilon}$ is the (now scalar) correlation between the single explanatory variable and the original model error term, ε.

In this case Equation (6) implies that

$$\sigma_\varepsilon^2 = \lambda' \Sigma_{XX}^{-1} \lambda + \sigma_\nu^2 = \frac{\sigma_{x\varepsilon}^2}{\sigma_x^2} + \sigma_\nu^2 = \frac{\rho_{x\varepsilon}^2 \sigma_x^2 \sigma_\varepsilon^2}{\sigma_x^2} + \sigma_\nu^2, \tag{8}$$

since λ is the scalar $\sigma_{x\varepsilon} = \rho_{x\varepsilon}\sqrt{\sigma_x^2 \sigma_\varepsilon^2}$ and Σ_{XX} is the scalar σ_x^2 in this section. Hence,

$$\sigma_\nu^2 = \left(1 - {\rho_{x\varepsilon}}^2\right) \sigma_\varepsilon^2. \tag{9}$$

This directly implies that

$$\sqrt{n}\left(\widehat{\beta}^{OLS} - \left(\beta + \frac{\sigma_{x\varepsilon}}{\sigma_x^2}\right)\right) \xrightarrow{D} N\left(0, \left(1 - \rho_{x\varepsilon}^2\right)\frac{\sigma_\varepsilon^2}{\sigma_x^2}\right). \tag{10}$$

as in Kiviet (2016, Equation 2.7) for the special case of a bivariate regression.

3. Our Proposed Sensitivity Analysis

In implementing the sensitivity analysis proposed here, we presume that the regression model equation given above as Equation (3) has been estimated using OLS, so that the sample data (realizations of Y and X) are available, and have been used to obtain a sample realization of the inconsistent OLS parameter estimator ($\widehat{\beta}_{OLS}$) – which is actually consistent for $\beta + \Sigma_{XX}^{-1}\lambda$ – and to obtain a sample realization of the usual estimator of the model error variance estimator, $s^2 = \widehat{\sigma}_\nu^2$. This error variance estimator provides a consistent estimate of

the Equation (3) error variance, σ_ν^2, conditional on the values of λ and Σ_{XX}. In addition, the sample length (n) is taken to be sufficiently large that $\sqrt{n}\,\widehat{\beta}_{OLS}$ has converged to its limiting distribution and that $\widehat{\sigma}_\nu^2$ has essentially converged to its probability limit, σ_ν^2. In the first part of this section it is also assumed that n is sufficiently large that the sample estimate $\widehat{\Sigma}_{XX}$ need not be distinguished from its probability limit, Σ_{XX}; this assumption is relaxed later in this section since – as Kiviet (2016) points out – sampling variance in $\widehat{\Sigma}_{XX}$ materially affects the asymptotic sampling distribution $\widehat{\beta}_{OLS}$, through its impact on the asymptotic bias in $\widehat{\beta}_{OLS}$.

Now assume – for a moment – that λ, the k-dimensional vector of covariances between the columns of the X matrix and the vector of errors (ε) in Equation (1) is posited and taken as given; this artificial assumption will be relaxed shortly. In that case the rejection p-value for any null hypothesis specifying a linear restriction on the components of the parameter vector β can be readily obtained, using the asymptotic sampling distribution given as Equation (4) above, leading to a test statistic distributed as Student's t with n - k degrees of freedom under this null hypothesis.[4]

A consistent estimator of β – call it $\widehat{\beta}_{consistent}$ – which, from Equations (3) and (4), is clearly just $\widehat{\beta}_{OLS} - \Sigma_{XX}^{-1}\lambda$ can then be easily obtained and substituted into Equation (1) to provide a set of model residuals which are asymptotically equivalent to the vector of structural model errors, ε. The sample variance of this implied ε vector then yields $\widehat{\sigma}_\varepsilon^2$, which is a consistent estimate of σ_ε^2, the variance of ε.[5]

[4]The rejection p-value for testing a null hypothesis instead specifying a set of $q > 1$ linear restrictions, to be jointly tested, similarly leads to a test statistic asymptotically distributed $F(q,\ n-k)$ and leads to a sensitivity analysis which is so similar that the exposition is couched here (solely for expositional clarity) in terms of a single linear restriction. For that matter, many econometrics programs - e.g., `Stata` - make it very easy to test a nonlinear restriction on the components of β; and the sensitivity analysis described here further extends to the (asymptotically valid) rejection p-values from such testing in a completely straightforward way.

[5]As noted above, this variance σ_ε^2 exceeds σ_ν^2, the fitting-error variance in the model as actually estimated using OLS, because the inconsistency in the OLS parameter estimate strips out of the estimated model errors the portions of ε which are - due to the assumed endogeneity - correlated with the columns of X. Note that $\sigma_\varepsilon^2 > \sigma_\nu^2$ also follows mathematically from Equation (6), because Σ_{XX} is positive definite.

This consistent estimate of the variance of ε is then combined with the posited λ covariance vector and with the consistent sample variance estimates for the k explanatory variables – i.e., with $\widehat{\Sigma}_{XX}(1,1)$, ..., $\widehat{\Sigma}_{XX}(k,k)$ – to yield $\widehat{\rho}_{X\varepsilon}$, a consistent estimate of the corresponding k-vector of correlations between these explanatory variables and the original model errors (ε) in Equation (1). As with the other sample quantities, we will assume that n is sufficiently large that the sampling errors in $\widehat{\rho}_{X\varepsilon}$ can be neglected. Thus, for any posited λ covariance vector, the concomitant correlation vector $\rho_{X\varepsilon}$ can effectively be readily calculated. This vector of correlations is worth estimating because it quantifies the (linear) endogeneity posited in each of the explanatory variables in a more intuitively interpretable way than does λ, the posited vector of covariances between the explanatory variables and the original model errors, itself. For the purpose of our sensitivity analysis we denote the Euclidean length of this implied correlation k-vector $\rho_{X\varepsilon}$ below as "$|\rho_{X\varepsilon}|$".

In summary, then, any posited λ vector of covariances between the explanatory variables and ε, the original model errors, yields an implied (asymptotically valid) rejection p-value for the null hypothesis at issue, and a consistent estimate of the vector $\rho_{X\varepsilon}$, whose components are the implied correlations between the k explanatory variables and the original model errors $\rho_{X\varepsilon}$, with Euclidean length $|\rho_{X\varepsilon}|$.

The value of the λ vector is, of course, unknown, so this calculation is repeated for a selection M of all possible values it can take on, retaining the aforementioned correlation vector ($\rho_{X\varepsilon}$), its length ($|\rho_{X\varepsilon}|$), and the concomitant null hypothesis rejection p-value. These values are then (for each chosen λ vector) written to one row of a spreadsheet file, if and only if the null hypothesis is no longer rejected at some designated p-value; here, for clarity of exposition only, this designated p-value is specified as 0.05. Because the regression model need not be re-estimated for each posited λ vector, these calculations are computationally inexpensive; consequently, it is quite feasible for M to range up to 10^5 or even 10^6.[6]

[6]The value of M is limited to more like $1,000$ to $10,000$ when (as described below) the sensitivity analysis is simulated multiple times so as to quantify the dispersion in its results generated by the likely sampling errors in $\widehat{\Sigma}_{XX}$.

For $\ell \leq 2$ – i.e., where the exogeneity of at most one or two of the k explanatory variables is taken to be suspect – it is computationally feasible to repeat the calculations using a straightforward ℓ-dimensional grid-search over the reasonably-possible λ vectors; for larger values of ℓ it is still feasible (and, in practice, effective for this purpose) to use a Monte-Carlo search instead, as described in Ashley and Parmeter (2015b), based on drawing λ vectors at random from a multi-dimensional Gaussian distribution.

The algorithm described above yields a spreadsheet containing $M' < M$ rows each containing an implied correlation k-vector $\rho_{X\varepsilon}$, its Euclidean length $|\rho_{X\varepsilon}|$, and its implied null hypothesis rejection p-value, with the latter quantity in each case, by construction, exceeding the nominal value of 0.05. For a sufficiently large value of M' this collection of $\rho_{X\varepsilon}$ vectors well-approximates an ℓ-dimensional set in the vector space spanned by the non-zero components of $\rho_{X\varepsilon}$. We denote this as the "No Longer Rejecting" or "NLR" set: the elements of this set are the X-column-to-ε correlations (exogeneity-assumption flaws) which are sufficient to overturn the 5%-significant null hypothesis rejection observed in the original OLS regression model.[7] Sorting this spreadsheet on the correlation-vector length $|\rho_{X\varepsilon}|$ then yields the point in the NLR which is closest to the origin – i.e., the smallest $\rho_{X\varepsilon}$ vector which represents a flaw in the exogeneity assumptions sufficient to overturn the observed rejection of the null hypothesis of interest at the 5% level. This vector – which we denote r_{min} – and its length, $|r|_{min}$, then quantify the sensitivity of this particular null hypothesis inference to possible endogeneity in any of these ℓ explanatory variables in the original regression model.

The computational burden of calculating r_{min} as described above – where the impact of the sampling errors in $\widehat{\Sigma}_{XX}$ is being neglected – is not large, so that `R` and `Stata` code (available from the authors) is generally quite sufficient to the task. But it is illuminating

[7]For simplicity of exposition this passage is written for the case where the null hypothesis is rejected in the original OLS model, so that λ vectors yielding p-values exceeding 0.05 are overturning this observed rejection. Where the original-model inference is a failure to reject the null hypothesis, then $\rho_{X\varepsilon}$, $|\rho_{X\varepsilon}|$, and the concomitant rejection p-value are instead written out to the spreadsheet file only when this p-value is less than 0.05. In this case one would instead denote this as the "No Longer Not Rejecting" set, but the sensitivity analysis is otherwise the same.

to obtain r_{min} analytically for the not-uncommon special case where $\ell = 1$ – i.e., where just one (the m^{th}, say) of the k explanatory variables is being taken as possibly-endogenous; in that case only the m^{th} component (λ_m) of the λ vector is non-zero.[8]

For the special case where $\ell = 1$, then – and (solely for expositional clarity) restricting attention to a sensitivity analysis with respect to the rejection p-value for a null hypothesis which corresponds to the particularly simple single linear restriction that $\beta_j = 0$ – it is easy to characterize the two values of λ_m for which the null hypothesis is barely rejected at the 5% level: For these two values of λ_m the concomitant bias induced in $\widehat{\beta}_{OLS}$ ($\Sigma_{XX}^{-1}\lambda$) must barely suffice to make the magnitude of the relevant estimated t ratio equal its 2.5% critical value, $t^c_{0.025}(n-k)$.

Thus, these two requisite λ_m values must each satisfy the equation

$$\left| \frac{b_j - \left(\widehat{\Sigma}_{XX}^{-1}(j,m)\right)\lambda_m}{\sqrt{s^2 \widehat{\Sigma}_{XX}^{-1}(j,j)}} \right| \;=\; t^c_{0.025}(n-k), \tag{11}$$

where b_j and s^2 are the sample realizations of $\widehat{\beta}_j$ and $\hat{\sigma}^2_\nu$ from OLS estimation of Equation (3) and $\widehat{\Sigma}_{XX}^{-1}(j,m)$ is the $(j,m)^{th}$ element of the inverse of $\widehat{\Sigma}_{XX}$, the (consistent) sample estimate of Σ_{XX}. Equation (11) generalizes in an obvious way to a null hypothesis which is a linear restriction on the components of β. And where heteroscedasticity and/or serial correlation issues are present in the model errors of Equation (3), one can simply replace the denominator in Equation (11) with the appropriate White-Eicker or HAC standard error estimate.

Equation (11) yields the two solution values,

$$\lambda_m^{\pm} \;=\; \frac{b_j \;\pm\; \sqrt{s^2 \left(\widehat{\Sigma}_{XX}^{-1}(j,j)\right)}\, t^c_{0.025}(n-k)}{\widehat{\Sigma}_{XX}^{-1}(j,m)} \,. \tag{12}$$

[8]Illumination aside, the substantial computational efficiency improvement afforded by this essentially analytic r_{min} calculation is quite useful when – below – a bootstrap-based simulation is introduced so as to obtain an estimated standard error for r_{min}, quantifying the dispersion induced in it when one allows for the likely sampling errors in $\widehat{\Sigma}_{XX}$.

Mathematically, there are two solutions to Equation (11) because of the absolute value function. Intuitively, there are two solutions because a larger value for λ_m increases the bias in the j^{th} component of $\widehat{\beta}_{OLS}$ at rate $\Sigma_{XX}^{-1}(j,m)$. Thus – supposing that this component of $\widehat{\beta}_{OLS}$ is (for example) positive – then sufficiently changing λ_m in one direction can reduce the value of $\widehat{\beta}_{consistent}$ just enough so that it remains positive and is now just barely significant at the 5% significance level; but changing λ_m sufficiently more in this direction will reduce the value of $\widehat{\beta}_{consistent}$ enough so that it becomes sufficiently negative as to again be barely significant at the 5% level.

These two values of λ_m lead to two implied values for the single non-zero (m^{th}) component of the implied correlation vector ($\rho_{X\varepsilon}$); r_{min} is then the one of these two vectors with the smallest magnitude, which magnitude is then $|r|_{min} = \min(|\lambda^-|, |\lambda^+|)$.

In summary ,then, this vector r_{min} is practical to calculate for any multiple regression model for which we suspect that one (or a number) of the explanatory variables might be endogenous to some degree: in the special case where there is a single such explanatory variable ($\ell = 1$) and the null hypothesis of interest is a single linear restriction on the β vector, then Equation (12) yields the analytical solution for r_{min} described immediately above. For $\ell > 1$ (or where the null hypothesis of interest is more complicated), the Monte Carlo search over M covariance (λ) vectors described above is easily programmed and is still computationally inexpensive. Either way, it is computationally straightforward to calculate r_{min} for any particular null hypothesis on the β vector; its length, $|r|_{min}$), then objectively quantifies the sensitivity of the rejection p-value for this particular null hypothesis to possible endogeneity in these ℓ explanatory variables.

But how, precisely, is one to interpret the value of this estimated value for $|r|_{min}$? Clearly, if $|r|_{min}$ is close to zero – less than around 0.10, say – then only a fairly small amount of explanatory-variable endogeneity suffices to invalidate the original OLS-model rejection of this particular null hypothesis at the 5% level. One could characterize such an inference as "fragile" with respect to possible endogeneity problems; and one might not want to place

much confidence in this null hypothesis rejection unless and until one is able to find credibly-valid instruments for the explanatory variables with regard to which inference is relatively fragile.[9] In contrast, a large value of $|r|_{min}$ – greater than around 0.40, say – indicates that quite a large amount of explanatory-variable endogeneity is necessary in order to invalidate the original OLS-model rejection of this null hypothesis at the 5% level. One could characterize such an inference as "robust" with respect to possible endogeneity problems, and perhaps not worry overmuch about looking for valid instruments in this case. Notably, inference with respect to one important and interesting null hypothesis might be fragile (or robust) with respect to possible endogeneity in one set of explanatory variables, whereas inference on another key null hypothesis might be differently fragile (or robust) – and with respect to possible endogeneity in a different set of explanatory variables: the sensitivity analysis results sensibly depend on the inferential question at issue.

But what about an intermediate estimated value of $|r|_{min}$? Such a result is indicative of an inference for which the issue of its sensitivity to possible endogeneity issues is still sensibly in doubt. Here the analysis again suggests that one should limit the degree of confidence placed in this null hypothesis rejection, unless and until one is able to find credibly-valid instruments for the explanatory variables which the sensitivity analysis indicates are problematic in terms of potential endogeneity issues. In this instance the sensitivity analysis has not clearly settled the fragility versus robustness issue, but at least it provides a quantification which is communicable to others, and which is objective in the sense that any analyst will obtain the same $|r|_{min}$ result. This situation is analogous to the ordinary hypothesis-testing predicament when a null hypothesis is rejected with a p-value of, say, 0.07: whether or not to reject the null hypothesis is not clearly resolved based on such a result, but one has at least objectively quantified the weight of the evidence against the null hypothesis.

[9]Perfectly exogenous instruments are generally unavailable also, so it is useful to note again that Ashley and Parmeter (2015b) provides an analogous sensitivity analysis procedure allowing one to quantify the robustness (or fragility) of IV-based inference rejection p-values to likely flaws in the instruments used.

In summary, here we have obtained the asymptotic distribution of the OLS parameter estimator for the general (k-variate) multiple regression model with linearly-endogenous regressors. This result allows us to analyze hypothesis test rejection p-value sensitivity in the usual multiple regression model to possible endogeneity in any combination of the explanatory variables, and for any specific null hypothesis restriction (or restrictions). Notably, we have obtained essentially analytic sensitivity analysis results for the special – but not particularly uncommon – situation where only one explanatory variable at a time is considered to be possibly-endogenous and the null hypothesis of interest is a single linear restriction on the coefficient vector, but we note that the numerical calculations required for the general case are not in fact computationally burdensome.

The foregoing discussion, however, has neglected the impact of the sampling errors in $\widehat{\Sigma}_{XX}$ on the asymptotic distribution of $\widehat{\beta}^{OLS}$, which was derived as Equation (4) above in terms of the population quantity Σ_{XX}. Substitution of $\widehat{\Sigma}_{XX}$ for Σ_{XX} in the Equation (4) asymptotic variance expression is asymptotically inconsequential, but Kiviet (2016) correctly points out that this substitution into the expression for asymptotic bias in Equation (4) has an impact on the asymptotic variance of $\widehat{\beta}^{OLS}$. Hence sampling errors in $\widehat{\Sigma}_{XX}$ will have an effect on the $|r|_{min}$ values calculated as described above. So as to gauge the magnitude of these effects, our implementing software estimates a standard error for each calculated $|r|_{min}$ value using row-wise bootstrap simulation based on the observed X matrix.

4. Revised Sensitivity Analysis Results for the Mankiw, Romer, and Weil (1992) Study of the Impact of Human Capital on Economic Growth

Ashley and Parmeter (2015a) provided sensitivity analysis results with regard to statistical inference on the two main hypotheses in the classic Mankiw, Romer, and Weil (MRW, 1992) study on economic growth: first, that human capital accumulation does impact growth, and

second that their main regression model coefficients sum to zero.[10] Here these results are updated using the corrected sampling distribution obtained in Section 1 in Equation (4).

For reasons of limited space the reader is referred to Ashley and Parmeter (2015a) – or MRW (1992) – for a more complete description of the MRW model. Here we will merely note that the dependent variable in the MRW regression model is real per capita GDP for a particular country and that the three MRW explanatory variables are the logarithm of the number of years of schooling ("$\ln(School)$", their measure of human capital accumulation), the real investment rate per unit of output ("$\ln(I/GDP)$"), and a catch-all variable ("$\ln(n+g+\delta)$"), capturing population growth, income growth, and depreciation).

Table 1 displays our revised sensitivity analysis results for both of the key MRW hypothesis tests considered in Ashley and Parmeter (2015a, Tables 1 and 2). The main changes here – in addition to using the corrected Equation (4) sampling distribution result – are that we now additionally include sensitivity analysis results allowing for possible exogeneity flaws in all three explanatory variables simultaneously, and that we now provide bootstrap-simulated standard error estimates which quantify the uncertainty in each of the $|r|_{min}$ values quoted due to sampling variation in $\widehat{\Sigma}_{XX}$.[11]

In Table 1 the analytic $|r|_{min}$ results – per Equations (11) and (12) – are used for the three columns in which a single explanatory variable is considered; the $|r|_{min}$ results simultaneously considering two or three explanatory variables were obtained using $M = 10,000$ Monte Carlo draws. The standard error estimates are in all cases based on 1,000 bootstrap simulations of $\widehat{\Sigma}_{XX}$. Table 1 quotes only the $|r|_{min}$ values and not the full ℓ-dimensional r_{min} vectors because these vectors did not clearly add to the interpretability of the results.[12]

[10] MRW (1992, page 421) explicitly indicates that this sum is to equal zero under the null hypothesis: the indication to the contrary in Ashley and Parmeter (2015a) was only a typographical error.

[11] We note that, because of the way his approach frames and tabulates the results, Kiviet's procedure cannot address scenarios in which either two or all three explanatory variables are simultaneously considered to be possibly endogenous.

[12] The full r_{min} vector is of much greater interest in the analogous sensitivity analysis with respect to the validity of the instruments in IV estimation/inference provided in Ashley and Parmeter (2015b). In that context, the relative fragility of the instruments is very much to the point, as one might well want to drop an instrument which leads to inferential fragility.

We first consider the $\ell = 1$ results for both null hypotheses, where the inferential sensitivity is examined with respect to possible endogeneity in one explanatory variable at a time. For these $\ell = 1$ results we find that the MRW inference result with respect to their rejection of $H_o : \beta_{school} = 0.0$ (at the 5% level) appears to be quite robust with respect to reasonably likely amounts of endogeneity in $\ln(n + g + \delta)$ and in $\ln(I/GDP)$, but not so clearly robust with respect to possible endogeneity in $\ln(School)$ once one takes into account the uncertainty in the $\ln(School)$ $|r|_{min}$ estimate due to likely sampling variation in $\widehat{\Sigma}_{XX}$.

TABLE 1. Sensitivity Analysis Results on $H_o : \beta_{school} = 0.0$ and $H_o : \beta_{school} + \beta_{I/GDP} + \beta_{ng\delta} = 0$ from Mankiw, Romer and Weil (1992). The "$\%|r|_{min} = 0$" row gives the percentage of bootstrap simulations yielding $|r|_{min}$ equal to zero for this null hypothesis.

Variable	$\ln(n + g + \delta)$	$\ln(I/GDP)$	$\ln(School)$	$\ln(I/GDP)$ & $\ln(School)$	All Three
			H_o: $\beta_{school} = 0.0$		
$\|r\|_{min}$	0.94	0.57	0.45	0.38	0.65
[*std.error*]	[0.03]	[0.09]	[0.08]	[0.06]	[0.11]
			H_o: $\beta_{school} + \beta_{I/GDP} + \beta_{ng\delta} = 0$		
$\|r\|_{min}$	0.11	0.22	0.72	0.28	0.59
[*std.error*]	[0.04]	[0.11]	[0.19]	[0.11]	[0.07]
$\%\|r\|_{min} = 0$	19.4 %	8.3 %	2.2%	2.0%	0.0%

In contrast, the MRW inference result with respect to their concomitant failure to reject $H_o : \beta_{school} + \beta_{I/GDP} + \beta_{ng\delta} = 0$ at the 5% level appears to be fairly fragile with respect to possible endogeneity in their $\ln(n + g + \delta)$ explanatory variable. In particular, our $|r|_{min}$ estimate is just 0.11 ± 0.04 for this case. This $|r|_{min}$ estimate is quite small – and $|r|_{min}$ is inherently non-negative – so one might naturally worry that its sampling distribution (due to sampling variation in $\widehat{\Sigma}_{XX}$) might be so non-normal that an estimated standard error could be misleading in this instance. We note, however, that $|r|_{min}$ is zero – because this MRW null hypothesis is actually rejected at the 5% level for any posited amount of endogeneity in $\ln(n + g + \delta)$ – in fully 19.4% of the bootstrapped $\widehat{\Sigma}_{XX}$ simulations. We consequently conclude that this $|r|_{min}$ estimate is indeed so small that there is quite a decent chance

of very modest amounts of endogeneity in $\ln(n+g+\delta)$ overturning the MRW inferential conclusion with regard to this null hypothesis.

With an $|r|_{min}$ estimate of 0.22 ± 0.11 (and $|r|_{min}$ turning up zero in 8.3% of the bootstrapped $\widehat{\Sigma}_{XX}$ simulations) there is also considerable evidence here that the MRW failure to reject H_o:$\beta_{school} + \beta_{I/GDP} + \beta_{ng\delta} = 0$ is fragile with respect to possible endogeneity in their $\ln(I/GDP)$ explanatory variable; but this result is a bit less compelling. The sensitivity analysis result for this null hypothesis with respect to possible endogeneity in the $\ln(School)$ variate is even less clear: the $|r|_{min}$ estimate (of 0.72) is large – indicating robustness to possible endogeneity, but this $|r|_{min}$ estimate comes with quite a substantial standard error estimate, indicating that this robustness result is not itself very stable across likely sampling variation in $\widehat{\Sigma}_{XX}$. These two results must be classified as "mixed."

The $\ell = 2$ and $\ell = 3$ results for these two null hypotheses are a bit difficult to interpret for this MRW data set: these results appear to be simply reflecting a mixture of the sensitivity results for the corresponding $\ell = 1$ cases, where each of the three explanatory variables is considered individually. These multidimensional sensitivity results are also computationally a good deal more burdensome to obtain, because they require Monte Carlo searches rather than following directly from the analytic results, as in Equations (11) and (12). It is in principle possible that a consideration of inferential sensitivity with respect to possible endogeneity in several explanatory variables at once (i.e. doing the sensitivity analysis with $\ell > 1$) might yield fragility-versus-robustness conclusions quite at variance with those obtained from the corresponding one-dimensional sensitivity analyses for some other data set. And there is no way to be certain that this is not the case without trying it out for those data; but our tentative conclusion is that the one-dimensional ($\ell = 1$) sensitivity analyses are of the greatest practical value.

In summary, then, this application to the MRW study provides a rather comprehensive look at what our sensitivity analysis can provide. First of all, we did not need to make any additional model assumptions beyond those already present in the MRW study. In

particular, our sensitivity analysis does not require the specification of any information with regard to the higher moments of any of the random variables – as in Kiviet(2016, 2018) – although it does restrict attention to linear endogeneity and brings in the issue of sampling variation in $\widehat{\Sigma}_{XX}$. Second, we were able to examine the robustness/fragility with respect to possible explanatory variable endogeneity for both of the key MRW inferences: one of which was the rejection of a simple zero-restriction and the other of which was the failure to reject a more complicated linear restriction.[13] Third, we found clear evidence of both robustness and fragility in the MRW inferences, depending on the null hypothesis and on the explanatory variables considered. Fourth, we did not – in this MRW data set – find that the multi-dimensional sensitivity analyses (with respect to two or three of the explanatory variables at a time) provided any additional insights not already clearly present in the one-dimensional sensitivity analysis results, the latter of which are computationally very inexpensive because they can use our analytic results. Finally, we find that the bootstrap-simulated standard errors in the $|r|_{min}$ values (arising due to sampling variation in the estimated explanatory variable variances) are already manageable – albeit not negligible – at the MRW sample length, of $n = 98$. Thus, doing the bootstrap simulations (already programmed in our implementing `R` and `Stata` software) is pretty clearly necessary with samples of this length; but it also sufficient, in that it suffices to yield useful sensitivity analysis results.[14]

5. Concluding Remarks

Because the 'structural' model errors – ε in Equation (1) – are unobservable, in empirical practice one cannot know (and cannot, even in principle, test) whether or not the explanatory

[13]In contrast, sensitivity analysis with respect to this linear restriction would have been awkward or impossible using the confidence-interval-centric approach proposed in Kiviet (2016). The joint null hypothesis that both of these linear restrictions hold could have been examined here also (either using the Monte Carlo algorithm or solving the quadratic polynomial equation which would result from the analog of Equation (11) in that instance), had that been deemed worthwhile in this instance.

[14]These standard errors would be smaller – and their estimation less necessary – in substantially larger samples. For substantially smaller samples these standard errors could make the sensitivity analysis noticeably less useful, but diagnostic checking is generically difficult in very small samples and regression analysis with stochastic regressors is in any case only justified asymptotically.

variables in a multiple regression model are or are not exogenous, without making additional (and untestable) assumptions. In contrast, we have shown here that it actually ***is*** possible to quantitatively investigate whether or not the rejection p-value (with regard to any specific null hypothesis that one is particularly interested in) either is (or is not) *sensitive* to likely amounts of linear endogeneity in the explanatory variables.[15]

This paper proposes a flexible procedure for engaging in just such an investigation, which can be implemented in actual practice, as a routine 'screen' for possible endogeneity problems in linear multiple regression estimation/inference settings. If this sensitivity analysis indicates that the inferences one most cares about are quite robust to likely amounts of correlation between the explanatory variables and the model errors, then one can defend one's use of OLS inference without further ado. If, in contrast, this sensitivity analysis shows that one's key inferences are fragile with respect to minor amounts of correlation between the explanatory variables and the model errors, then a serious consideration of more sophisticated estimation approaches – such as those proposed in Caner and Morrill (2013), Kraay (2012), Lewbel (2012), or Kiviet (2018) – is both warranted and motivated. Where the sensitivity analysis is indicative of an intermediate level of robustness/fragility in the inference, then one at least has an objective indication as to the predicament that one is in.

Finally, we note that Kiviet (2016) suggests a different way of displaying sensitivity analysis results so as to quantify the potential impact of unmodeled endogeneity on OLS inference: his suggestion is to tabulate the 95% confidence intervals (for the components of the structural coefficient vector which are of greatest interest) against assumed values for $\rho_{X\varepsilon}$, the vector of correlations between this explanatory variables and the structural model error. This approach is quite attractive where one's interest centers on the coefficient values themselves rather than on inference with respect to null hypotheses specifying one or more linear

[15]The restriction to 'linear endogeneity' made here dramatically simplifies the analysis: linear endogeneity in an explanatory variable means that it is related to the unobserved model error, but solely in a linear fashion. The correlations of the explanatory variables with the model error completely capture the endogeneity relationship if and only if the endogeneity is linear; thus – where the endogeneity contemplated is in any case going to be expressed in terms of such correlations – little is lost in restricting attention to linear endogeneity.

restrictions on them; it becomes awkward or infeasible, however, where $\ell \geq 2$ – i.e., where one is allowing for possible exogeneity in more than a single explanatory variable at a time – as one would then need to examine $(\ell + 1)$-dimensional tables. Still, we can readily envision settings where the exogeneity of just a single explanatory variable is problematic and where one's interest centers on the sensitivity of structural coefficient confidence intervals rather than on the sensitivity of tests on linear restrictions with regard to the structural coefficients; we see Kiviet's sensitivity-results display approach as preferable in such settings.

References

[1] Ashley, R., and Parmeter, C. F. 2015a. When is it Justifiable to Ignore Explanatory Variable Endogeneity in a Regression Model?", *Economics Letters*, 137(1), 70-74.

[2] Ashley, R., and Parmeter, C. F. 2015b. "Sensitivity Analysis for Inference in 2SLS Estimation with Possibly-Flawed Instruments," *Empirical Economics*, 49(4), 1153-1171.

[3] Caner, M. and Morrill, M. S., 2013. "Violation of Exogeneity: A Joint Test of Structural Parameters and Correlation," Working paper.

[4] Davidson, R. and MacKinnon, J. G. 1993. *Estimation and Inference in Econometrics*, Oxford University Press, New York New York.

[5] Friedman, M. 1953. *Essays in Positive Economics*, University of Chicago Press, Chicago, IL.

[6] Johnston, J. 1992. *Econometric Methods*, McGraw-Hill, New York New York.

[7] Kinal, T. W. 1980. "The Existence of Moments of k-Class Estimators," *Econometrica*, 48, 241-249.

[8] Kiviet, J. F. 2013. "Identification and Inference in a Simultaneous Equation Under Alternative Information Sets and Sampling Schemes," *The Econometrics Journal*, 16, S24-S59.

[9] Kiviet, J. F. 2016. "When is it Really Justifiable to Ignore Explanatory Variable Endogeneity in a Regression Model?," *Economics Letters*, 145, 192-195.

[10] Kiviet, J. F. 2018. "Testing the Impossible: Identifying Exclusion Restrictions," Working Paper.

[11] Kiviet, J. F. and Niemczyk, J. 2007. "The Asymptotic and Finite Sample Distributions of OLS and Simple IV in Simultaneous Equations," *Computational Statistics & Data Analysis*, 51, 3296-3318.

[12] Kiviet, J. F. and Niemczyk, J. 2012. "Comparing the Asymptotic and Empirical (Un)conditional Distributions of OLS and IV in a Linear Static Simultaneous Equation," *Computational Statistics & Data Analysis*, 56, 3567-3586.

[13] Kraay, A. 2012. "Instrumental Variables Regression with Uncertain Exclusion Restrictions: A Bayesian Approach," *Journal of Applied Econometrics*, 27, 108-128.

[14] Lewbel, A. 2012. " Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models," *Journal of Business and Economic Statistics*, 30(1), 67-80.

[15] Mankiw, N. G., Romer, D. and Weil, D. N. 1992. "A Contribution to the Empirics of Economic Growth," *The Quarterly Journal of Economics*, 107(2) pp. 407-437.

[16] Spanos, A. 1999. *Probability Theory and Statistical Inference*, Cambridge University Press, Cambridge.