SENSITIVITY ANALYSIS OF OLS MULTIPLE REGRESSION INFERENCE WITH RESPECT TO POSSIBLE LINEAR ENDOGENEITY IN THE EXPLANATORY VARIABLES, FOR BOTH MODEST AND FOR EXTREMELY LARGE SAMPLES

RICHARD A. ASHLEY AND CHRISTOPHER F. PARMETER

ABSTRACT. This work describes a versatile and readily-deployable sensitivity analysis of OLS inference with respect to possible endogeneity in the explanatory variables of the usual k-variate linear multiple regression model. This sensitivity analysis is based on a derivation of the sampling distribution of the OLS parameter estimator, extended to the setting where some, or all, of the explanatory variables are endogenous. In exchange for restricting attention to possible endogeneity which is solely linear in nature – the most typical case – no additional model assumptions must be made, beyond the usual ones for a model with stochastic regressors. The sensitivity analysis quantifies the sensitivity of hypothesis test rejection P-values and/or estimated confidence intervals to such endogeneity, enabling an informed judgement as to whether any selected inference is 'robust' versus 'fragile'. The usefulness of this sensitivity analysis – as a 'screen' for potential endogeneity issues – is illustrated with an example from the empirical growth literature. This example is extended to an extremely large sample, so as to illustrate how this sensitivity analysis can be applied to parameter confidence intervals in the context of massive datasets, as in 'Big Data'.

JEL Classification: C2, C15.

1. INTRODUCTION

1.1. **Motivation.** 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 to obtain consistent OLS parameter estimation and asymptotically valid statistical

inference. 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 (2009) and Ashley and Parmeter (2015b), which address similar issues in the context of IV and GMM/2SLS 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 acceptable to make possibly-false (and even very-likely-false) assumptions – *if and only if one can and does show that the model results one cares most about are insensitive to the levels of violations in these assumptions that it is reasonable to expect.*[1]

In particular, the present paper proposes a sensitivity analysis for OLS estimation/inference in the presence of unmodeled endogeneity in the explanatory variables of the usual linear multiple regression model. This context provides an ideal setting in which to both exhibit and operationalize a quantification of the 'insensitivity' alluded to in the proposition above, 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 common in applied economic work. The extension of this kind of sensitivity analysis to more complex – e.g., nonlinear – estimation settings is feasible, but will be laid out in separate work, as it requires a different computational framework, and does not admit the closed-form results obtained (for some cases) here.

The diagnostic analysis proposed here quantifies the sensitivity of OLS hypothesis test rejection P-values (for both multiple and/or nonlinear parameter restrictions) – and also of single-parameter confidence intervals – with respect to possible endogeneity (i.e. non-zero correlations between the explanatory variables and the model errors) in the usual k-variate multiple regression model.

---

[1]See Leamer (1983) for an early recognition of this corollary.

This sensitivity analysis rests on a derivation of the sampling distribution of $\widehat{\beta}^{OLS}$, the OLS parameter estimator, extended to the case where some or all of the explanatory variables are endogenous to a specified degree. In exchange for restricting attention to possible endogeneity which is solely linear in nature – the most typical case – the derivation of this sampling distribution proceeds in a particularly straightforward way. And under this 'linear endogeneity' restriction, no additional model assumptions – e.g. with respect to the 3rd and 4th moments of the joint distribution of the explanatory variables and the model errors – are necessary, beyond the usual assumptions made for any model with stochastic regressors. The resulting analysis quantifies the sensitivity of hypothesis test rejection P-values (and/or estimated confidence intervals) to such linear endogeneity, enabling the analyst to make an informed judgement as to whether any selected inference is 'robust' versus 'fragile' with respect to likely amounts of endogeneity in the explanatory variables.

We show below that, in the context of the linear multiple regression model, this additional sensitivity analysis is so straightforward an addendum to the usual OLS diagnostic checking which applied economists are already doing – both theoretically, and in terms of the effort required to set up and use our proposed procedure – that analysts can routinely use it. In this regard we see our sensitivity analysis as a general diagnostic 'screen' for possibly-important unaddressed endogeneity issues.

1.2. **Relation to Previous Work on OLS Multiple Regression Estimation/Inference.**
Deficiencies in previous work on the sensitivity of OLS multiple regression estimation/inference with regard to possible endogeneity in the explanatory variables – comprising Ashley and Parmeter (2015a) and Kiviet(2016) – has stimulated us to revise and extend our earlier work on this topic. What's new here is:

- **Improved Sampling Distribution Derivation, via a Restriction to 'Linear Endogeneity'**

In response to the critique in Kiviet (2016) of the OLS parameter sampling distribution under endogeneity given in Ashley and Parmeter (2015a), we now obtain this asymptotic sampling distribution, in a particularly straightforward fashion, by limiting the universe of possible endogeneity to solely-linear relationships between the explanatory variables and the model error. In Section 2 we show that the OLS parameter sampling distribution does not then depend on the 3rd and 4th moments of the joint distribution of the explanatory variables and the model errors under this linear-endogeneity restriction. This improvement on the Kiviet and Niemczyk (2007, 2012) and Kiviet (2013, 2016) sampling distribution results – which make no such restriction on the form of endogeneity, but do depend on these higher moments – is crucial to an empirically-usable sensitivity analysis: we see no practical value in a sensitivity analysis which quantifies the impact of untestable exogeneity assumptions only under empirically-inaccessible assumptions with regard to these higher moments. On the other hand, while endogeneity that is linear in form is an intuitively understandable notion – and corresponds to the kind of endogeneity that will ordinarily arise from the sources of endogeneity usually invoked in textbook discussions – this restriction to the sensitivity analysis is itself somewhat limiting; this issue is further discussed in Section 2 below.

- **Analytic Results for a Non-Trivial Special Case**

    We now obtain closed-form results for the minimal degree of explanatory-variable to model-error correlation necessary in order to overturn any particular hypothesis testing result for a single linear restriction, in the special (one-dimensional) case where the exogeneity of a single explanatory variable is under scrutiny.[2]

---

[2]Our results are easily-computable (but not analytic) for the general (multi-dimensional) sensitivity analysis proposed here – where possible (linear) endogeneity in any number of explanatory variables is under consideration, and where what's at issue is either a confidence interval for a single parameter or the null hypothesis rejection P-value for a test of multiple linear (and/or non-linear) restrictions: in that general setting our R and Stata implementations then require a bit more in the way of numerical computations. These issues are addressed in Sections 3 and 5 below.

- **Simulation-Based Check with Regard to Sample-Length Adequacy**

  Because the asymptotic bias in $\widehat{\beta}^{OLS}$ depends on Σ_{XX}, the population variance-covariance matrix of the explanatory variables in the regression model, sampling error in $\widehat{\Sigma}_{XX}$ – the usual (consistent) estimator of Σ_{XX} – does affect the asymptotic distribution of $\widehat{\beta}^{OLS}$ when this estimator replaces Σ_{XX} in our sensitivity analysis, as pointed out in Kiviet (2016). Our algorithm now optionally uses bootstrap simulation to quantify the impact of this replacement on our sensitivity analysis results.

  We find this impact to be noticeable, but manageable, in our illustrative empirical example, the classic Mankiw, Romer, and Weil (MRW, 1992) examination of the impact of human capital accumulation ("$\ln(School)$"), which uses $n = 98$ observations. This example does not necessarily imply that this particular sample length is either necessary or sufficient for the application of this sensitivity analysis to other models and data sets; we consequently recommend performing this simulation-based check (on the adequacy of the sample length for this purpose) at the outset of the sensitivity analysis in any particular regression modeling context.[3]

- **Sensitivity Analysis with Respect to Coefficient Size – Especially for Extremely Large Samples**

  In some settings – financial and agricultural economics, for example – analysts' interest centers on the ***size*** of a particular model coefficient rather than on the P-value at which some particular null hypothesis with regard to the model coefficients can be rejected. The estimated size of a model coefficient is generally also a better guide to its economic (as opposed to its statistical) significance – as emphasized by McCloskey and Ziliak (1996).

---

[3]These simulations do require a non-negligible computational effort, but this calculation needs only to be done once in any particular setting, and is already coded up in the `Stata` implementation of the algorithm. One might argue that a more wide-ranging initial simulation effort (subsuming this particular calculation) has value in any case, as a check on the degree to which the sample is more broadly sufficiently large for the use of the usual asymptotic inference machinery.

Our sensitivity analysis adapts readily in such cases, to instead depict how the estimated 95% confidence interval for a particular coefficient varies with the vector of correlations between the explanatory variables and the model error term; we denote this is as 'parameter-value-centric' sensitivity analysis below. For a one-dimensional analysis, where potential endogeneity is contemplated in only a single explanatory variable, the display of such results is easily embodied in a plot of the estimated confidence interval for this specified regression coefficient versus the single endogeneity-correlation in this setting.

We supply two such plots, embodying two such one-dimensional 'parameter-value-centric' sensitivity analyses, in Section 4 below; these plots display how the estimated 95% confidence interval for the MRW human capital coefficient (on their "$\ln(School)$" explanatory variable in our illustrative example) separately varies with the posited endogeneity correlation in each of two selected explanatory variables in the model. Our implementing algorithm itself generalizes easily to multi-dimensional sensitivity analyses – i.e., sensitivity analysis with respect to possible (simultaneous) endogeneity in more than one explanatory variable – and several examples of such two-dimensional sensitivity analysis results, with regard to the robustness/fragility of the rejection P-values for hypothesis tests involving coefficients in the MRW model, are presented in Section 4. We could thus additionally display the results of a two-dimensional 'parameter-value-centric' sensitivity analysis for this model coefficient – e.g., plotting how the estimated confidence interval for the MRW human capital coefficient varies with respect to possible simultaneous endogeneity in both of these two MRW explanatory variables – but the display of these confidence interval results would require a three-dimensional plot, which is more troublesome to interpret.[4]

---

[4]The underlying sensitivity analysis itself extends quite easily to a consideration of more than two possibly-endogeneous explanatory variables, but the resulting still-higher-dimensional confidence interval plots of this nature would be outright infeasible to visualize, however. On the other hand, one could in principle resort to tabulating the confidence interval results from such multi-dimensional 'parameter-value-centric' sensitivity analyses; Kiviet (2016) has already suggested such a tabulation as a sensitivity analysis display mechanism.

Perhaps the most interesting and important applications of this 'parameter-value-centric' sensitivity analysis will arise in the context of models estimated using the extremely large data sets now colloquially referred to as 'Big Data'. In these models, inference in the form of hypothesis test rejection P-values is frequently of limited value, because the large sample sizes in such settings quite often render all such P-values uninformatively small. In contrast, the 'parameter-value-centric' version of our endogeneity sensitivity analysis adapts gracefully to such large-sample applications: in these settings each estimated confidence interval will simply shrink to what amounts to a single point, so in such cases the 'parameter-value-centric' sensitivity analysis will in essence be simply plotting the asymptotic mean of a key coefficient estimate against posited endogeneity-correlation values. It is, of course, trivial to artificially extend our MRW example to a huge sample and display such a plot for illustrative purposes, and this is done in Section 4 below. But consequential examples of this sort of sensitivity analysis must await the extension of the OLS-inference sensitivity analysis proposed here to the nonlinear estimation procedures typically used in such large-sample empirical work.[5]

1.3. **Preview of the Rest of the Paper.** In Section 2 we provide a new derivation of the asymptotic sampling distribution of the OLS structural parameter estimator $(\widehat{\beta}^{OLS})$ for the usual k-variate multiple regression model, where some or all of the explanatory variables are endogenous to a specified degree. This degree of endogeneity is quantified by a given set of covariances between these explanatory variables and the model error term (ε); these covariances are denoted by the vector λ. The derivation of the sampling distribution of $\widehat{\beta}^{OLS}$ is greatly simplified by restricting the form of this endogeneity to be solely linear in form, as follows:

---

[5]Examples of such work include Rasmussen, et al. (2011) on stroke mortality in $N = 57,053$ Danish patients and Arellano, et al. (2018) on microeconometric consumption expenditures modeling using the U.S. PSID and the extensive Norwegian population-register administrative data set.

We first define a new regression model error term (denoted ν), from which all of the linear dependence on the explanatory variables has been stripped out. Thus – under this restriction of the endogeneity to be purely linear in form – this new error term must (by construction) be completely unrelated to the explanatory variables. Hence, with solely-linear endogeneity, ν must be statistically independent of the explanatory variables. This allows us to easily construct a modified regression equation in which the explanatory variables are independent of the model errors. This modified regression model now satisfies the assumptions of the usual multiple regression model (with stochastic regressors that are independent of the model errors), for which the OLS parameter estimator asymptotic sampling distribution is well known.

Thus, under the linear-endogeneity restriction, OLS estimation of this modified regression model yields unbiased estimation and asymptotically valid inferences on the model's parameter vector; but these OLS estimates are now, as one might expect, unbiased for a coefficient vector which differs from β by an amount depending explicitly on the posited endogeneity covariance vector, λ. Notably, however, this sampling distribution derivation requires no additional model assumptions beyond the usual ones made for a model with stochastic regressors: in particular, one need not specify any 3rd or 4th moments for the joint distribution of the model errors and explanatory variables, as would be the case if the endogeneity were not restricted to be solely-linear in form.

In Section 3 we show how this sampling distribution for $\widehat{\beta}^{OLS}$ can be used to assess the robustness/fragility of an OLS regression model inference with respect to possible endogeneity in the explanatory variables; this Section splits the discussion into several parts. The first parts of this discussion – Sections 3.1 through Section 3.5 – describe the sensitivity analysis with respect to how much endogeneity is required in order to 'overturn' an inference result with respect to the testing of a particular null hypothesis regarding the structural parameter, β, where this null hypothesis has been rejected at some nominal significance level – e.g., 5% – under the assumption that all of the explanatory variables are exogenous. Section 3.6 closes

our sensitivity analysis algorithm specification by describing how the sampling distribution for $\widehat{\beta}^{OLS}$ derived in Section 2 can be used to display the sensitivity of a confidence interval for a particular component of β to possible endogeneity in the explanatory variables; this is the 'parameter-value-centric' sensitivity analysis outlined at the end of Section 1.2. So as to provide a clear 'road-map' for this Section, each of these six subsections is next briefly described.

Section 3.1 shows how a specific value for the posited endogeneity covariance vector (λ) – combined with the sampling distribution of $\widehat{\beta}^{OLS}$ from Section 2 – can be used to both re-compute the rejection P-value for the specified null hypothesis with regard to β and to also convert this endogeneity covariance vector (λ) into the corresponding endogeneity correlation vector, $\hat{\rho}_{X\varepsilon}$, which is more-interpretable than λ. This conversion into a correlation vector is possible (for any given value of λ) because the $\widehat{\beta}^{OLS}$ sampling distribution yields a consistent estimator of β, making the error term ε in the original (structural) model asymptotically available; this allows the necessary variance of ε to be consistently estimated.

Section 3.2 operationalizes these two results from Section 3.1 into an algorithm for a sensitivity analysis with regard to the impact (on the rejection of this specified null hypothesis) of possible endogeneity in a specified subset of the k regression model explanatory variables; this algorithm calculates a vector we denote as 'r_{min}', whose Euclidean length – denoted '$|r|_{min}$' here – is our basic measure of the robustness/fragility of this particular OLS regression model hypothesis testing inference.

The definition of r_{min} is straightforward: it is simply the shortest endogeneity correlation vector, $\hat{\rho}_{X\varepsilon}$, for which possible endogeneity in this subset of the explanatory variables suffices to raise the rejection P-value for this particular null hypothesis beyond the specified nominal level (e.g., 0.05) – thus overturning the null hypothesis rejection observed under the assumption of exogenous model regressors. Since the sampling distribution derived in Section 2 is expressed in terms of the endogeneity covariance vector (λ), this implicit search proceeds in the space of the possible values of λ, using the P-value result from Section 3.1

to eliminate all λ values still yielding a rejection of the null hypothesis, and using the $\hat{\rho}_{X\varepsilon}$ result corresponding to each non-eliminated λ value to supply the relevant minimand, $|\hat{\rho}_{X\varepsilon}|$.

In Section 3.3 we obtain a closed-form expression for r_{min} in the (not uncommon) special case of a one-dimensional sensitivity analysis, where only a single explanatory variable is considered possibly-endogenous. This closed-form result reduces the computational burden involved in calculating r_{min}, by eliminating the numerical minimization over the possible endogeneity covariance vectors; but its primary value is didactic: its derivation illuminates what is going on in the calculation of r_{min}.

The calculation of r_{min} is not ordinarily computationally burdensome, even for the general case of multiple possibly-endogenous explanatory variables and tests of complex null hypotheses – with the sole exception of the simulation calculations alluded to in Section 1.2 above. These bootstrap simulations, quantifying the impact of substituting $\widehat{\Sigma}_{XX}$ for Σ_{XX} in the r_{min} calculation, are detailed in Section 3.4. In practice these simulations are quite easy to do, as they are already coded up as an option in the implementing software, but the computational burden imposed by the requisite set of $N_{boot} \approx 1000$ replications of the r_{min} calculation can be substantial. In the case of a one-dimensional sensitivity analysis – where the closed-form results are available – these analytic results dramatically reduce the computational time needed for this simulation-based assessment of the extent to which the length of the sample data set is sufficient to support the sensitivity analysis. And our sensitivity results for the illustrative empirical example examined in Section 4 suggest that such one-dimensional sensitivity analyses may frequently suffice. Still, it is fortunate that this simulation-based assessment in practice needs only to be done once, at the outset of one's work with a particular regression model.

The portion of Section 3 describing the sensitivity analysis with respect to inference in the form of hypothesis testing then concludes with some preliminary remarks, in Section 3.5, as to how one can interpret r_{min} (and its length, $|r|_{min}$) in terms of the 'fragility' or 'robustness'

of such a hypothesis test inference. This topic is taken up again in Section 6, at the end of the paper.

Section 3 closes with a description of the implementation of the 'parameter-value-centric' sensitivity analysis outlined at the end of Section 1.2. This version of the sensitivity analysis is simply a display of how the 95% confidence interval for any particular component of β varies with $\hat{\rho}_{X\varepsilon}$. Its implementation is consequently a straightforward variation on the algorithm for implementing the hypothesis-testing-centric sensitivity analysis, as the latter already obtains both the sampling distribution of $\widehat{\beta}^{OLS}$ and the value of $\hat{\rho}_{X\varepsilon}$ for any given value of the endogeneity covariance vector, λ. Thus, each value of λ chosen yields a point in the requisite display. The results of a univariate sensitivity analysis (where only one explanatory variable is considered to be possibly endogenous, and hence only one component of λ can be non-zero) are readily displayed in a two-dimensional plot. Multi-dimensional 'parameter-value-centric' sensitivity analysis results are also easy to compute, but these are more challenging to display; in such settings one could, however, still resort to a tabulation of the results, as in Kiviet (2016).

In Section 4 we illustrate the application of this proposed sensitivity analysis to assess the robustness/fragility of the inferences obtained by Mankiw, Romer, and Weil (1992) – denoted 'MRW' below – in their influential study of the impact of human capital accumulation on economic growth. This model provides a nice example of the application of our proposed sensitivity analysis procedure, in that the exogeneity of several of their explanatory variables is in doubt, and in that one of their two key null hypotheses is a simple zero-restriction on a single model coefficient and the other is a linear restriction on several of their model coefficients. We find that some of their inferences are robust with respect to possible endogeneity in their explanatory variables, whereas others are fragile. Hypothesis testing was the focus in the MRW study, but this setting also allows us to display how a 95% confidence interval for one of their model coefficients varies with the degree of (linear) endogeneity in a

selected explanatory variable, both for the actual MRW sample length ($n = 98$) and for an artificially-huge elaboration of their data set.

We view the sensitivity analysis proposed here as a practical addendum to the profession's usual toolkit of diagnostic checking techniques for OLS multiple regression analysis – in this case as a 'screen' for assessing the impact of likely amounts of endogeneity in the explanatory variables. To that end – as noted above – we have written scripts encoding the technique for several popular computing frameworks: `R` and `Stata`; Section 5 briefly discusses exactly what is involved in using these scripts.

Finally, in Section 6 we close the paper with a brief summary and a modest elaboration of our Section 3.5 comments on how to interpret the quantitative (objective) sensitivity results which this proposed technique provides.

2. Sampling Distribution of the OLS Multiple Regression Parameter Estimator With Linearly-Endogenous Covariates

Per the usual k-variate multiple regression modeling framework, we 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.

We next define a random variate ν which is – by construction – uncorrelated with X:

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

where it is readily verified that both $E(\nu)$ and $E(\frac{1}{n}X'\nu)$ equal zero,[6] and where we let var(ν)= $\sigma_\nu^2 I$ define the scalar σ_ν^2.

Solving Equation (2) for ε and substituting into Equation (1) yields

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

This is the regression model that one is actually estimating when one regresses Y on X, yielding a realization of $\widehat{\beta}^{OLS}$. Since the error term (ν) in this equation is constructed to be uncorrelated with X, $\widehat{\beta}^{OLS}$ is a consistent estimator, but for the actual coefficient vector in this model – $\beta + \Sigma_{XX}^{-1}\lambda$ – rather than for β. The reader will doubtless recognize the foregoing as the usual textbook derivation that OLS yields an inconsistent estimator of β in the presence of endogeneity in the model regressors.

What is needed here, however, is the full asymptotic sampling distribution for $\widehat{\beta}^{OLS}$, so that rejection P-values for null hypotheses involving β and confidence intervals for components of β can be computed for non-zero posited amounts of endogeneity in the original model, Equation (1). Kiviet and Niemczyk (2007, 2012) and Kiviet (2013, 2016) derive this sampling distribution; but their result, unfortunately, depends on the 3rd and 4th moments of the joint distribution of (X, ε). Absent consistent estimation of β, the Equation (1) model error (ε) is not observable, so these moments are not empirically accessible; those sampling distribution results are consequently not useful for implementing our sensitivity analysis.

Instead, we here limit the possible endogeneity considered in the sensitivity analysis to purely linear relationships between the explanatory variables (X) and the Equation (1) model errors, ε. Under the restriction of *solely linear* endogeneity, the error term ν is not just uncorrelated with X: it is actually *independent* of X. This result follows directly from the fact that Equation 2 defines ν as the remainder of the variation in ε once its linear

---

[6]Multiplication of Equation (2) by $n^{-1}X'$ and taking expectations yields $E(n^{-1}X'\nu) = E(n^{-1}X'\varepsilon - n^{-1}X'X\Sigma_{XX}^{-1}\lambda) = \lambda - \lambda = 0$.

relationship with X has been removed. Thus, if the relationship between ε and X is purely linear, then ν and X must be completely unrelated – i.e., statistically independent.

With X restricted to be linearly endogenous, the implied error term (ν) in the rightmost portion of Equation (3) is consequently independent of the model regressors. So this Equation (3) model is now reduced to the standard stochastic-regression model with (exogenous) regressors that are independent of the model error term, the only difference being that $\widehat{\beta}^{OLS}$ is here an unbiased estimator of $\beta + \Sigma_{XX}^{-1}\lambda$ rather than of β. The asymptotic sampling distribution of $\widehat{\beta}^{OLS}$ is thus well known:[7]

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

and this result makes it plain that $\widehat{\beta}^{OLS}$ is itself unbiased and consistent if and only if $\lambda = 0$.

The value of σ_ν^2 can thus, under this solely-linear endogeneity restriction, be consistently estimated from the OLS fitting errors – using the usual estimator, s^2.[8] And the value of Σ_{XX}^{-1} in Equation (4) can be consistently estimated from the observed data on the explanatory variables, using the usual sample estimator of the variance-covariance matrix of these variables. We note, however, that – per the discussion in Section 1 above – sampling variation in this estimator of Σ_{XX} impacts the distribution of $\widehat{\beta}^{OLS}$ – even asymptotically – through its influence on the estimated asymptotic mean implied by Equation (4); this complication is addressed in the implementing algorithm detailed in Section 3 below, using bootstrap simulation.

The restriction here to the special case of solely-linear endogeneity thus makes the derivation of the sampling distribution of $\widehat{\beta}^{OLS}$ almost trivial. But it is legitimate to wonder what this restriction is approximating away, relative to instead – as in Kiviet (2016), say – making an *ad hoc* – and empirically inaccessible – assumption as to the values of the 3rd and 4th

---

[7]E.g., from Johnston, 1972 Section 9-2, or from many earlier sources.

[8]The asymptotic variance of $\widehat{\beta}^{OLS}$ given by Equation (4) would no longer be valid where $\text{var}(\nu) \neq \sigma_\nu^2 I$, but one could in that case use White-Eicker or Newey-West standard error estimates.

moments of the joint distribution of the variables (X, ε). As noted in Section 1.2 above, the forms of endogeneity usually discussed in textbook treatments of endogeneity – i.e., endogeneity due to omitted explanatory variables, due to measurement errors in the explanatory variables, and due to a (linear) system of simultaneous equations – in fact all lead to endogeneity which is solely-linear in form. But endogeneity in a particular explanatory variable which is arising from its determination as the dependent variable of another equation in a system of nonlinear simultaneous equations will in general not be linear in form. So this restriction is clearly not an empty one. Moreover – in simulating data from a linear regression model – it would be easy to generate nonlinear endogeneity in any particular explanatory variable, by simply adding a nonlinear function of this explanatory variable to the model errors driving the simulations.[9]

Lastly, we note that Ashley and Parmeter (2019, p. 12) shows that the sampling distribution of $\widehat{\beta}^{OLS}$ derived above as Equation 4 is equivalent to that given by Kiviet (2016, Equation 2.7) for the special case where the joint distribution of (X, ε) is symmetric with zero excess kurtosis. Thus, joint Gaussianity in (X, ε) – which, of course, implies linear endogeneity – is sufficient, albeit not necessary, in order to make the two distributional results coincide. As noted above in Section 1.2, however, we see no practical value in a sensitivity analysis which quantifies the impact of untestable exogeneity assumptions on X and ε only under untestable assumptions with regard to the higher moments of the joint distribution of (X, ε). Thus, we would characterize the limitation of this sensitivity analysis to solely-linear endogeneity as a practically necessary – and relatively innocuous – but not empty, restriction.

Finally, it will be useful below to have an expression for σ_{ε}^2, the variance of the error term in Equation (1), so as to provide for the conversion of any posited value for λ, the k-vector of

---

[9]Repeatedly simulating data sets on (Y, X, ε) in this way would lead to a simulated distribution for (X, ε) which could not be jointly Gaussian and it would lead to a collection of simulated $\widehat{\beta}^{OLS}$ values whose variance would in fact differ as one varied the value of λ used in the simulation generating mechanism, in the manner predicted by the sampling distribution result given in Kiviet (2016).

covariances of X with ε, into the more-easily-interpretable k-vector of correlations between X and ε, which will be denoted $\rho_{X\varepsilon}$ below.

It follows from Equation (2) that

$$\varepsilon'\varepsilon = \left(X\Sigma_{XX}^{-1}\lambda + \nu\right)' \left(X\Sigma_{XX}^{-1}\lambda + \nu\right)$$

$$(5) \qquad = \lambda'\Sigma_{XX}^{-1}X'X\Sigma_{XX}^{-1}\lambda + 2\lambda'\Sigma_{XX}^{-1}X'\nu + \nu'\nu.$$

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

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

It makes sense that σ_ε^2 exceeds σ_ν^2: 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 – in the presence of endogeneity – Equation (3) can use sample variation in some of its included variables to 'explain' some of the sample variation in ε. 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 ($\widehat{\beta}^{OLS}$) is then inconsistent for the Equation (1) structural parameter, β.

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 identifiable. It *is*, however, possible – quite easy, actually – to quantify the *sensitivity* of the rejection P-value for any particular null hypothesis involving β (or of a 95% confidence interval for any particular component of β) to correlations between the explanatory variables and these unobservable structural errors. The next section describes

our algorithm for accomplishing this sensitivity analysis, which is the central result of the present paper.

3. Proposed Sensitivity Analysis Algorithm

3.1. Obtaining Inference Results and the Endogeneity Correlation Vector $(\widehat{\rho}_{X\varepsilon})$, Given the Endogeneity Covariance Vector (λ). 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 the sampling distribution of $\widehat{\beta}_{OLS}$ has converged to its limiting distribution – derived above as Equation (4) – and that $\widehat{\sigma}_\nu^2$ has essentially converged to its probability limit, σ_ν^2. In the first parts 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 for the bootstrap simulations described below in Section 3.4.

Now assume – for the moment – that a value for 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 is taken as given; this artificial assumption will be relaxed shortly. Here (in an "ℓ-dimensional sensitivity analysis") ℓ of the explanatory variables are taken to be possibly-endogenous, and hence ℓ components of this posited λ vector would typically be non-zero.

We presume that the inference at issue – whose sensitivity to possible endogeneity in the model explanatory variables is being assessed – is a particular null hypothesis specifying a given set of $r \geq 1$ linear restrictions on the components of β. Based on this posited λ vector –

and the sampling distribution for $\widehat{\beta}^{OLS}$, derived as Equation (4) above – the usual derivation then yields a test statistic which is asymptotically distributed as $F(r, n - k)$ under this null hypothesis. Thus, the P-value at which this null hypothesis of interest can be rejected is readily calculated, for any given value of the posited λ vector.[10]

Equations (3) and (4) together imply a consistent estimator of β: $\widehat{\beta}^{consistent} = \widehat{\beta}_{OLS} - \Sigma_{XX}^{-1}\lambda$. So substitution of this estimator into Equation (1) yields 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.[11]

This consistent estimate of the variance of ε is then combined with the posited covariance vector (λ) and with the consistent sample variance estimates for the k explanatory variables – i.e., with $\widehat{\Sigma}_{XX}(1,1), \ldots, \widehat{\Sigma}_{XX}(k,k)$ – 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).

Thus, the concomitant correlation vector $\widehat{\rho}_{X\varepsilon}$ can be readily calculated for any posited covariance vector, λ. This k-dimensional 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. We denote the Euclidean length of this implied correlation vector $\widehat{\rho}_{X\varepsilon}$ below as "$|\widehat{\rho}_{X\varepsilon}|$".

In summary, then, any posited value for λ – the vector of covariances between the explanatory variables and ε, the original model errors – yields both an implied rejection P-value for

---

[10]Many econometrics programs – e.g., `Stata` – make it equally easy to calculate analogous rejection P-values for null hypotheses which consist of a specified set of nonlinear restriction on the components of β; so the `Stata`-based implementation of the sensitivity analysis described here readily extends to tests of nonlinear null hypotheses.

[11]As shown in Section 2, 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.

the null hypothesis at issue and also $\hat{\rho}_{X\varepsilon}$, a consistent estimate of the vector of implied correlations between the k explanatory variables and the original model errors, with Euclidean length $|\hat{\rho}_{X\varepsilon}|$.

3.2. Calculation of r_{min}, the Minimal-Length Endogeneity Correlation Vector Overturning the Hypothesis Test Rejection.

The value of the λ vector is, of course, unknown, so the calculation described above is repeated for a selection M of all possible values λ can take on, so as to numerically determine the λ vector which yields the smallest value of the endogeneity correlation vector length ($|\hat{\rho}_{X\varepsilon}|$) for which the null hypothesis of interest is no longer rejected at the nominal significance level – here taken (for clarity of exposition only) to be 5% – that is already being used. This minimal-length endogeneity correlation vector is denoted by r_{min}, and its length by $|r|_{min}$. The value of $|r|_{min}$ then quantifies the sensitivity of this particular null hypothesis inference to possible endogeneity in any of these ℓ explanatory variables in the original regression model.

Thus, for each of the M λ vectors selected, the calculation retains the aforementioned correlation vector ($\hat{\rho}_{X\varepsilon}$), its length ($|\hat{\rho}_{X\varepsilon}|$), and the concomitant null hypothesis rejection P-value; these values are then written to one row of a spreadsheet file, if and only if the null hypothesis is no longer rejected with P-value less than 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.[12]

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 select the M λ vectors 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 instead use a

---

[12]The value of M is limited to more like $1,000$ to $10,000$ when (as described in Section 3.4) the entire sensitivity analysis is simulated multiple times so as to quantify the dispersion in $|r|_{min}$ generated by the likely sampling errors in $\hat{\Sigma}_{XX}$.

Monte-Carlo search over the set of reasonably-possible λ vectors, as described in Ashley and Parmeter (2015b); in this case the λ vectors are drawn at random.[13]

The algorithm described above yields a spreadsheet containing M' rows, each containing an implied correlation k-vector $\hat{\rho}_{X\varepsilon}$, its Euclidean length $|\hat{\rho}_{X\varepsilon}|$, and its implied null hypothesis rejection P-value – with the latter quantity in each case exceeding the nominal rejection criterion value (e.g., 0.05), by construction. For a sufficiently large value of M', this collection of $\hat{\rho}_{X\varepsilon}$ vectors well-approximates an ℓ-dimensional set in the vector space spanned by the ℓ non-zero components of the vector $\hat{\rho}_{X\varepsilon}$. 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.[14]

Simply sorting this spreadsheet on the correlation-vector length $|\hat{\rho}_{X\varepsilon}|$ then yields the point in the NLR which is closest to the origin – i.e., r_{min}, the smallest $\hat{\rho}_{X\varepsilon}$ 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.

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 – in Section 3.3 below – to obtain r_{min} analytically for the not-uncommon special case where

---

[13]The resulting value of $|r|_{min}$ is itself insensitive to the form of the distribution used to generate the λ vectors used in such a Monte-Carlo search, so long as this distribution has sufficient dispersion. A multivariate Gaussian distribution was used in Ashley and Parmeter (2015b), but the reader is cautioned that this is a numerical device only: Gaussianity is not being assumed thereby for any random variable in the econometric analysis.

[14]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 instead a failure to reject the null hypothesis, then $\hat{\rho}_{X\varepsilon}$, $|\hat{\rho}_{X\varepsilon}|$, 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 collection of $\hat{\rho}_{X\varepsilon}$ vectors as the 'No Longer Not Rejecting' set; the sensitivity analysis is otherwise the same.

$\ell = 1$ – i.e., where just one (the mth, say) of the k explanatory variables is being taken as possibly-endogenous; in that case only the mth component (λ_m) of the λ vector is non-zero.[15]

3.3. Closed-Form Result for the Minimal-Length Endogeneity Correlation (r_{min}) Overturning the Hypothesis Test Rejection in the One-Dimensional Case.

In this section r_{min} is obtained in closed form for the special – but not uncommon – case of a one-dimensional $(\ell = 1)$ sensitivity analysis, where the impact on the inference of interest is being examined with respect to possible endogeneity in a single one – the mth – of the k explanatory variables in the regression model. The explicit derivation given here is for a sensitivity analysis with respect to the rejection of the simple null hypothesis that $\beta_j = 0$, but this restriction is solely for expositional clarity: an extension of this result to the sensitivity analysis of a null hypothesis specifying a given linear restriction – or set of linear restrictions – on the components of β is not difficult, and is indicated in the text below.

In the analysis of $H_0 : \beta_j = 0$ 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 asymptotic bias induced in $\widehat{\beta}_{OLS}$ – i.e., $\Sigma_{XX}^{-1}\lambda_m$ – must barely suffice to make the magnitude of the relevant estimated t ratio equal its 2.5% critical value, $t_{0.025}^c(n - k)$.

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

(7)
$$\left| \frac{b_j - \left(\widehat{\Sigma}_{XX}^{-1}(j, m) \right) \lambda_m}{\sqrt{s^2 \widehat{\Sigma}_{XX}^{-1}(j, j)}} \right| = t_{0.025}^c(n - k),$$

where b_j and s^2 are the sample realizations of $\widehat{\beta}_j$ and $\hat{\sigma}_\nu^2$ 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 (7) generalizes in an obvious way to a null hypothesis which is a linear restriction on the components of β; this extension merely makes the notation more

---

[15]Illumination aside, the substantial computational efficiency improvement afforded by this essentially analytic r_{min} calculation is quite useful when – in Section 3.4 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 $\widehat{\Sigma}_{XX}$.

complicated; for a set of r linear restrictions, the test statistic is – in the usual way – squared and the requisite critical value is then the 0.050 fractile of the $F(r, n - k)$ distribution.[16]

Equation (7) yields the two solution values,

$$(8) \qquad \lambda_m^{\pm} = \frac{b_j \pm \sqrt{s^2 \left(\widehat{\Sigma}_{XX}^{-1}(j, j)\right)} t_{0.025}^c(n - k)}{\widehat{\Sigma}_{XX}^{-1}(j, m)} .$$

Mathematically, there are two solutions to Equation (7) because of the absolute value function. Intuitively, there are two solutions because a larger value for λ_m increases the asymptotic bias in the jth 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 $\widehat{\beta}_j^{consistent} = b_j - \left(\widehat{\Sigma}_{XX}^{-1}(j, m)\right) \lambda_m$ 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 $\widehat{\beta}_j^{consistent}$ 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 $(\hat{\rho}_{X\varepsilon})$; r_{min} is then the one of these two vectors with the smallest magnitude, which magnitude is then:

$$(9) \qquad |r|_{min} = \frac{\min(|\lambda_m^-|, |\lambda_m^+|)}{\sqrt{\widehat{\sigma}_\varepsilon^2 \, \widehat{\Sigma}_{XX}(m, m)}}$$

where $\widehat{\sigma}_\varepsilon^2$ the consistent estimate of the variance of ε defined (along with $\widehat{\beta}^{consistent}$) in Section 3.1 above.

3.4. Simulation-Based Check on Sample-Length Adequacy with Regard to Substituting $\widehat{\Sigma}_{XX}$ for Σ_{XX}.

The foregoing discussion 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)

---

[16]Where heteroscedasticity and/or serial correlation issues are present in the model errors of Equation (3), the appropriate White-Eicker or HAC standard error estimate can be used in place of the denominator in Equation (7).

in Section 2 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 the asymptotic mean in Equation (4) has an impact on the distribution of $\widehat{\beta}^{OLS}$ which is not asymptotically negligible. Sampling errors in $\widehat{\Sigma}_{XX}$ could hence, in principle, have a notable effect on the $|r|_{min}$ values calculated as described above.

So as to gauge the magnitude of these effects on this measure of inference robustness/fragility, our implementing software estimates a standard error for each calculated $|r|_{min}$ value, using bootstrap simulation based on the observed X matrix. The size of this estimated standard error for $|r|_{min}$ is in effect quantifying the degree to which the sample length (n) is sufficiently large as to support the kind of sensitivity analysis proposed here.[17]

This standard error calculation is a fairly straightforward application of the bootstrap, which has been amply described in an extensive literature elsewhere, but a few comments are warranted here.[18]

First, it is essential to simulate new X matrix data 'row-wise' – that is choosing the ith row of the newly-simulated X matrix as the jth row of the original X data matrix, where j is chosen randomly from the integers in the interval $[1, n]$ – so as to preserve any correlations between the k explanatory variables in the original data set.

Second, it is important to note that the ordinary (Efron) bootstrap is crucially assuming that the rows of the original X matrix are realizations of IID – which is to say, independently and identically distributed – vectors. The 'identically distributed' part of this assumption is essentially already 'baked in' to our sensitivity analysis (and into the usual results for stochastic regressions like Equation 1), at least for the variances of the explanatory variables:

---

[17]We note that, where n is not adequate to the estimation of Σ_{XX} in a particular regression setting, then skepticism is also warranted in this setting as to whether the size of n is more broadly inadequate for asymptotic parameter inference work – i.e., estimation of confidence intervals and/or hypothesis test rejection P-values – even aside from concerns with regard to explanatory variable endogeneity.

[18]Davidson and MacKinnon (1993, pp. 764-5) provides a brief description of the mechanics of the usual bootstrap and a list of references to this literature, including (of course) Efron (1979).

otherwise the diagonal elements of Σ_{XX} are not constants for which the corresponding diagonal elements of $\widehat{\Sigma}_{XX}$ could possibly provide consistent estimates. Where the observations in Equation 1 are actually in time-order, then the 'independently distributed' part of the IID assumption can be problematic: the bootstrap simulation framework is in that setting assuming that each explanatory variable is serially independent, which is frequently not the case with economic time-series data. In this case, however, one could (for really large samples) bootstrap block-wise; or one could readily estimate a k-dimensional VAR model for the rows of the X matrix and bootstrap-simulate out of the resulting k-vector residuals, which are at least (by construction) serially uncorrelated.

That said, it is easy enough in practice to use the bootstrap to simulate $N_{boot} \approx 1,000$ replications of X, of $\widehat{\Sigma}_{XX}$, and hence of $|r|_{min}$. The standard deviation of these N_{boot} replicants of $|r|_{min}$ is not intended as a means to make precise statistical inferences as to the (population) value of $|r|_{min}$, but it is adequate to meaningfully gauge the degree to which n is sufficiently large that it reasonable to use the $|r|_{min}$ values obtained using the original data set in this kind of sensitivity analysis.

The calculation of a single r_{min} estimate (quantifying the sensitivity analysis of the rejection P-value a particular null hypothesis – even a complicated one – to possible endogeneity in a particular set of explanatory variables) is ordinarily not computationally burdensome. Repeating this calculation N_{boot} times can become time-consuming, however, unless the sensitivity analysis is one-dimensional – i.e., is with respect to possible endogeneity in a single explanatory variable – in which case the closed form results obtained in Section 3.3 above speed up the computations tremendously. For this reason, we envision this simulation-based standard error calculation as something one typically does only occasionally – e.g., at the outset of one's sensitivity analysis work on a particular regression model, as a check on the adequacy of the sample length to this purpose.

3.5. **Preliminary Remarks on the Interpretation of the Length of r_{min} as Characterizing the Robustness/Fragility of the Inference.** The vector r_{min} is thus 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 ($\ell = 1$), then the discussion in Section 3.3 – including the explicit result in Equation (9) for the special case where the null hypothesis of interest is a single linear restriction on the β vector – yields a closed-form solution for r_{min}. For $\ell > 1$, the search over M endogeneity covariance (λ) vectors described in Section 3.2 – so as to calculate r_{min} as the closest point to the origin in the 'No Longer Rejecting' set – 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 user-chosen significance level. One could characterize the statistical inference based on such a rejection as 'fragile' with respect to possible endogeneity problems; and one might not want to place much confidence in this inference unless and until one is able to find credibly-valid instruments for the explanatory variables with regard to which inference is relatively fragile.[19]

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 at the user-chosen significance

---

[19]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.

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 thus 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 particular 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.

3.6. Extension to 'Parameter-Value-Centric' Sensitivity Analysis.

As noted at the end of Section 1.2, in some settings the thrust of the analysis centers on the **size** of a particular component of β rather than on the P-value at which some particular null hypothesis with regard to the components of β can be rejected.

In such instances the sensitivity analysis proposed here actually simplifies: in such a 'parameter-value-centric' sensitivity analysis one merely needs to display how the estimated

95% confidence interval for this particular coefficient varies with $\hat{\rho}_{X\varepsilon}$, the posited implied vector of endogeneity correlations between the explanatory variables and the model error term.

This variation on the sensitivity analysis described in Sections 3.1 through 3.5 above is computationally much easier, since it is no longer necessary to choose the endogeneity co-variance vector (λ) so as to minimize the length of r_{min}. Instead – for a selected set of n_{plot} λ vectors – one merely exploits the $\widehat{\beta}_{OLS}$ sampling distribution derived in Section 2 to both estimate a 95% confidence interval for this particular component of β and (per Section 3.1) to convert this value for λ into the concomitant endogeneity correlation vector, $\hat{\rho}_{X\varepsilon}$.[20] The 'parameter-value-centric' sensitivity analysis then simply consists of either plotting or tabu-lating these n_{plot} estimated confidence intervals against the corresponding values calculated for $\hat{\rho}_{X\varepsilon}$.

Such a plot is an easily-displayed two-dimensional figure for a one-dimensional sensitivity analysis – i.e., for $\ell = 1$, where only a single explanatory variable is considered to be possibly-endogenous. For a two-dimensional sensitivity analysis – where $\ell = 2$ and two explanatory variables are taken to be possibly-endogenous – this plot requires the display of a three-dimensional figure; this is feasible, but it is more troublesome to plot and interpret. A graphical display of this kind of sensitivity analysis is infeasible for larger values of ℓ, but one could still tabulate the results.[21]

The most intriguing aspect of this 'parameter-value-centric' approach will arise in the sensitivity analysis – with respect to possible endogeneity issues – of 'Big Data' models, using huge data sets. The P-values for rejecting most null hypotheses in these models are so small as to be uninformative, whereas the 95% confidence for the key parameter estimates

---

[20]A straightforward rearrangement of Equation 7 yields the 95% confidence interval for the jth component of β, where endogeneity is considered possible in the mth explanatory variable and the value of the single non-zero component in the endogeneity covariance vector is λ_m.

[21]Such a tabulation was suggested in Kiviet (2016) as a way to display the kind of sensitivity analysis results proposed in Ashley and Parmeter (2015a); his sampling distribution for $\widehat{\beta}_{OLS}$, as highlighted prior, requires knowledge of the 3rd and 4th moments of the joint distribution of the explanatory variables and the model errors, however.

in these models are frequently especially informative because their lengths shrink to become tiny. Thus, the 'parameter-value-centric' sensitivity analysis is in this case essentially plotting how the estimated parameters vary with the posited amount of endogeneity correlation. This kind of analysis is potentially quite useful, as it is well-known that there is no guarantee whatever that the explanatory variables in these models are particularly immune to endogeneity problems.

We provide several examples of this kind of sensitivity analysis at the end of the next section, where the applicability and usefulness both kinds of sensitivity analysis are illustrated using a well-known empirical model from the literature on determinants of economic growth. In particular, the 'parameter-value-centric' sensitivity analysis results for this model are artificially extended to a huge data set at the end of Section 4 by the simple artifice of noting that the confidence-intervals plots displayed in the 'parameter-value-centric' sensitivity analysis shrink in length, essentially to single points, as the sample length becomes extremely large. However, more interesting and consequential examples of the application of 'parameter-value-centric' sensitivity analysis to 'Big Data' models must await the extension of the OLS sensitivity analysis proposed here to the nonlinear estimation procedures typically used in such work, as in the Rasmussen et al. (2011) and Arellano, et al. (2018) studies cited at the very end of Section 1.2.

4. Revised Sensitivity Analysis Results for Mankiw, Romer, and Weil (1992)

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.[22] Here these results are updated using the corrected sampling distribution obtained in Section 1 in Equation (4).

---

[22]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.

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 $\widehat{\Sigma}_{XX}$.[23]

In Table 1 the analytic $|r|_{min}$ results – per Equations (7) and (8) – 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.[24]

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

---

[23]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 more explanatory variables are simultaneously considered to be possibly endogenous.

[24]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.

$H_0 : \beta_{school} = 0.0$ (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 $H_0 : \beta_{school} = 0.0$ and $H_0 : \beta_{school} + \beta_{I/GDP} + \beta_{ng\delta} = 0$ 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_0: $\beta_{school} = 0.0$		
$|r|_{min}$	0.94	0.57	0.45	0.38	0.65
$[std.error]$	[0.03]	[0.09]	[0.08]	[0.06]	[0.11]
			H_0: $\beta_{school} + \beta_{I/GDP} + \beta_{ng\delta} = 0$		
$|r|_{min}$	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 $H_0 : \beta_{school} + \beta_{I/GDP} + \beta_{ng\delta} = 0$ 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 boot-strapped $\widehat{\Sigma}_{XX}$ simulations) there is also considerable evidence here that the MRW failure to

reject $H_0 : \beta_{school} + \beta_{I/GDP} + \beta_{ng\delta} = 0$ 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 not as easy 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. And these multidimensional sensitivity results are also computationally a good deal more burdensome to obtain, because they require Monte Carlo or grid searches, rather than following directly from the analytic results in Equations (8) and (9). While it is in principle possible that a consideration of inferential sensitivity with respect to possible endogeneity in several explanatory variables at once might yield fragility-versus-robustness conclusions at variance with those obtained from the corresponding one-dimensional sensitivity analyses for some other data set, we think this quite unlikely for the application of this kind of sensitivity analysis to linear regression models. Consequently, our tentative conclusion is that the one-dimensional ($\ell = 1$) sensitivity analyses are of the greatest practical value here.

In summary, then, this application of the 'hypothesis-testing-centric' sensitivity analysis to the MRW study provides a rather comprehensive look at what this aspect of 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.[25]

- 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 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.

- And, fifth, we determined 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 is pretty clearly necessary with a sample of this length; but it also sufficient, in that it suffices to show that this sample is sufficiently long as to yield useful sensitivity analysis results.[26]

Finally, we apply the 'parameter-value-centric' sensitivity analysis described in Section 3.6 to the MRW study. Their paper is fundamentally about the impact of human capital

---

[25]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 (7) in that instance), had that been deemed worthwhile in this instance.

[26]These standard error estimates would be smaller – and their estimation less necessary – in substantially larger samples. In substantially smaller samples, however, these bootstrap-simulated standard errors could in principle be so large as to make the sensitivity analysis noticeably less useful; it is consequently valuable that this addendum to the sensitivity analysis allows one to easily check for this. We envision a practitioner routinely obtaining these bootstrap-simulated standard errors for the $|r|_{min}$ values. However – because of its computational cost – we see these calculations being done only once or twice: at the outset of the analysis for any particular model and data set, and perhaps again at the very end.

– quantified by the years of schooling variable, "ln($School$)" – on output; so testing the null hypothesis $H_0 : \beta_{school} = 0.0$ was a crucial concern for them in terms of economic theory. Perhaps more relevant for economic policy-making, however, might be the size of their estimated coefficient on ln($School$), $\hat{\beta}_{school}^{OLS}$. Their estimated model yields a 95% confidence interval of [0.510, 0.799] for β_{school}, but this estimated confidence interval contains β_{school} in repeated samples with probability 0.95 only if their model assumptions are all valid. These notably include the assumptions that their most important explanatory variables – ln(I/GDP) and ln($School$) – are both exogenous.

As described in Section 3.6, we calculated how this OLS-estimated 95% confidence interval for β_{school} varies with possible endogeneity in each of ln($School$) and ln(I/GDP), and present the results of these one-dimensional 'parameter-value-centric' sensitivity analyses below as Figures 1 and 2. Figure 1 graphs how the estimated 95% confidence interval for β_{school} varies with the concomitant correlation between the model errors and ln(I/GDP), both of which are jointly generated over a selection of 1,000 values for the covariance between the model errors and the ln(I/GDP) variable. Figure 2 analogously graphs how the estimated 95% confidence interval for β_{school} varies with the concomitant correlation between the model errors and the ln($School$) variable. This correlation, plotted on the horizontal axis, is in each figure the single non-zero component of the endogeneity correlation vector, $\hat{\rho}_{X\varepsilon}$.

We note that:

- We could just as easily have used 500 or 5,000 endogeneity-covariance values: because (in contrast to the computation of $|r|_{min}$) these calculations do not involve a numerical minimization, the production of these figures requires only a few seconds of computer time regardless.
- These plots of the upper and lower limits of the confidence intervals are nonlinear – although not markedly so – and the (vertical) length of each interval does depend (somewhat) on the degree of endogeneity-correlation.

FIGURE 1. Estimated 95% confidence interval for β_{school} graphed against posited correlation between the model errors and $\ln(I/GDP)$.
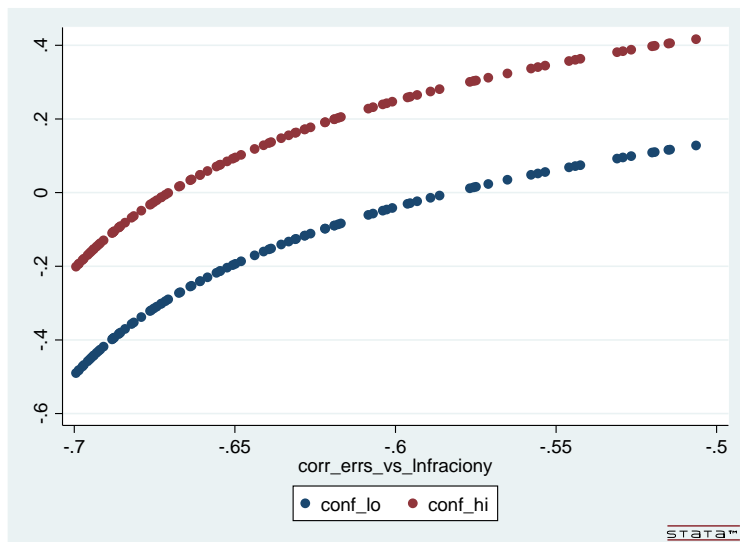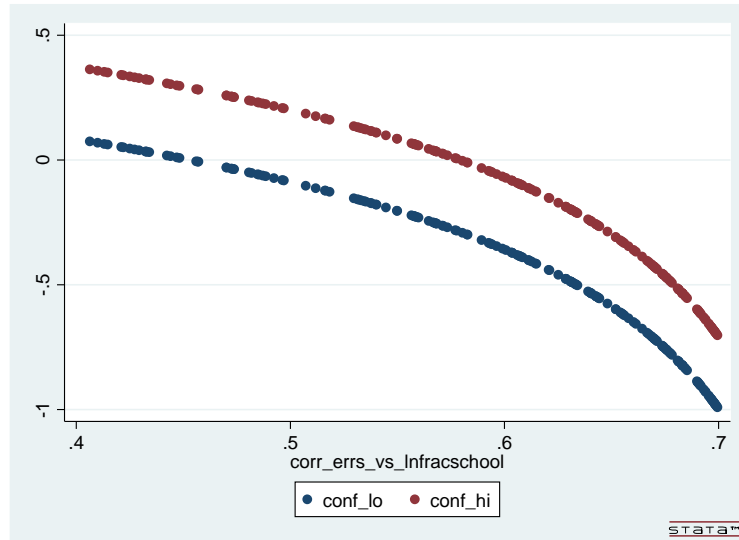
FIGURE 2. Estimated 95% confidence interval for β_{school} graphed against posited correlation between the model errors and $\ln(School)$.

- For a two-dimensional 'parameter-value-centric' sensitivity analysis – analogous to the "$\ln(I/GDP)$ & $\ln(School)$" column in Table 1 listing our 'hypothesis-testing-centric' sensitivity analysis results – the result would be a single three-dimensional plot, displaying (as its height above the horizontal plane) how the 95% confidence interval for β_{school} varies with the two components of the endogeneity-correlation vector which are now non-zero. In the present (linear multiple regression model) setting, this three-dimensional plot is graphically and computationally feasible, but not clearly more informative than Figures 1 and 2 from the pair of one-dimensional analyses. And for a three-dimensional 'parameter-value-centric' sensitivity analysis – e.g., analogous to the "All Three" column in Table 1 – it would still be easily feasible to compute and tabulate how the 95% confidence interval for β_{school} varies with the three components of the endogeneity-correlation vector which are now non-zero; but the resulting four dimensional plot is neither graphically renderable nor humanly visualizable, and a tabulation of these confidence intervals (while feasible to compute and print out) is not readily interpretable.

- One can easily read off from each of these two figures the result of the corresponding one-dimensional 'hypothesis-testing-centric' sensitivity analysis for the null hypothesis $H_0 : \beta_{school} = 0.0$ – i.e., the value of $|r|_{min}$: One simply observes the magnitude of the endogeneity-correlation value for which the graph of the upper confidence interval limit crosses the horizontal ($\beta_{school} = 0$) axis and the magnitude of the endogeneity-correlation value for which the graph of the lower confidence interval limit crosses this axis: $|r|_{min}$ is the smaller of these two magnitudes.[27]

- Finally, if the MRW sample were expanded to be huge – e.g., by repeating the 98 observations on each variable an exceedingly large number times – then the upper and lower limits of the 95% confidence intervals plotted in Figures 1 and 2 would collapse into one another. The plotted confidence intervals in these figures would

---

[27]These two crossings correspond to the two solutions for λ_m in Equation (8) in Section 3.3.

then essentially become line plots, and each figure would then simply be a plot of the 'ln($School$)' component of $\widehat{\beta}^{consistent}$ versus the the single non-zero component of the endogeneity correlation vector, $\hat{\rho}_{X\varepsilon}$.[28] Such an expansion is, of course, only notional for an actual data set of modest length, but – in contrast to hypothesis testing in general (and hence to 'hypothesis-testing-centric' sensitivity analysis in particular – this feature of the 'parameter-value-centric' sensitivity analysis results illustrates how useful this kind of analysis would remain in the context of modeling the huge data sets recently becoming available, where the estimational/inferential distortions arising due to unaddressed endogeneity issues do not diminish as the sample length expands.

5. Implementation Considerations

The sensitivity analysis procedure proposed here is 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, only a small barrier to its adoption and use. Importantly, the analysis requires no additional assumptions with regard to the regression model at issue, beyond the ones which would ordinarily need to be made in any OLS multiple regression analysis with stochastic regressors. In particular, this sensitivity analysis does not require the user to make any additional assumptions with regard to the higher moments of the joint distribution of the explanatory variables and the model errors.

For most empirical economists the computational implementation of this sensitivity analysis is very straightforward: since `Stata` and `R` scripts implementing the algorithms are available, these codes can be simply patched into the script that the analyst is likely already using for estimating their model. All that the user needs to provide, then, is a choice as to which of the explanatory variables which are under consideration in the sensitivity analysis

---

[28]As derived in Section 3.1 $\widehat{\beta}^{consistent} = \widehat{\beta}_{OLS} - \Sigma_{XX}^{-1}\lambda$, the consistent estimator of β, given the endogeneity covariance vector, λ.

as being possibly-endogenous, and a specification of the particular null hypothesis (or of the particular 95% parameter confidence interval) whose robustness or fragility with respect to this possible endogeneity is of interest.[29]

Our implementing routines can readily handle the 'hypothesis-testing-centric' sensitivity analysis with respect to any sort of null hypothesis (simple or compound, linear or nonlinear) for which `Stata` and `R` can compute a rejection P-value, using whatever kind of standard error estimates (OLS/White-Eicker/HAC) are already being used. For this kind of sensitivity analysis these routines calculate $|r|_{min}$, where r_{min} is defined in Section 1.3 above 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. These implementing scripts can also, optionally, perform the set of bootstrap simulations described in Section 3.4 above, so as to additionally estimate standard errors for the $|r|_{min}$ estimates – reflecting the impact of substituting the sample variance-covariance matrix of the explanatory variables $(\widehat{\Sigma}_{XX})$ for its population value (Σ_{XX}) in the Equation (4), as described in Section 3.4.)[30]Alternatively, where the 'parameter-value-centric' sensitivity analysis (with respect to a confidence interval for a particular model coefficient) is needed, the implementing software tabulates this estimated confidence interval versus a selection of $\hat{\rho}_{X\varepsilon}$ values, for use in making a plot.

_____

[29]For completeness, we note that the user also needs to provide a value for M which is sufficiently large that the routine can numerically find the shortest endogeneity-correlation vector reaching the 'No Longer Rejecting' set with adequate accuracy and, if needed, the number of bootstrap simulations to be used in estimating standard errors for the $|r|_{min}$ estimates.

[30]Three versions of the software are available. One version implements the analytic result derived in Section 3.3 for the value of λ minimizing the length of $\hat{\rho}_{X\varepsilon}$; this version is written for the special case of a null hypothesis which is a single linear restriction and a single possibly-endogenous explanatory variable, but could be fairly easily extended to the testing of more complicated null hypotheses. A second version already allows for the analysis of any kind of null hypothesis – even compound or nonlinear – but still restricts the dimensionality of the sensitivity analysis to possible endogeneity in a single explanatory variable; this script requires a line-search over putative values of λ – here scalar-valued – but is still very fast. A third version extends the sensitivity analysis to the simultaneous consideration of possible endogoeneity in multiple explanatory variables; this version is a bit more computationally demanding, as it implements either a multidimensional grid search or a Monte-Carlo minimization over a range of λ vectors.

6. Concluding Remarks

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

Our approach does restrict the sensitivity analysis to possible endogeneity which is solely linear in form: this restriction to 'linear endogeneity' dramatically simplifies the analysis in Sections 2 and 3. 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. Moreover – as noted in Section 2 above – all of the usually-discussed sources of endogeneity are, in fact, linear in this sense.[31] And the meaning of this restriction is so clearly understandable that the user is in a position to judge its restrictiveness for themselves. 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, so it seems to us quite unlikely that a sensitivity analysis limited to solely-linear possible endogeneity will to any substantial degree underestimate the fragility of one's OLS inferences with respect to likely amounts of endogeneity.

---

[31]The joint determination of the dependent variable and one or more of the explanatory variables in a set of *nonlinear* simultaneous equations would yield nonlinear endogeneity, but this is not the kind of simultaneity most usually discussed.

This paper proposes a flexible procedure for engaging in just such a sensitivity analysis – and one which is so easily implemented in actual practice that it can reasonably serve as a routine 'screen' for possible endogeneity problems in linear multiple regression estimation/inference settings.[32] This screen produces an objective measure – the calculated 'r_{min}' vector and its Euclidean length '$|r|_{min}$' – which quantify the sensitivity of the rejection of a chosen, particular null hypothesis (one that presumably is of salient economic interest) to possible endogeneity in the vector of k explanatory variables in the linear multiple regression model of Equation (1). The relative sizes of the components of this r_{min} vector yield an indication as to which explanatory variables are relatively problematic in terms of endogeneity-related distortion of this particular inference; and this length $|r|_{min}$ yields an overall indication as to how 'robust' versus 'fragile' this particular inference is with respect to possible endogeneity in the explanatory variables considered in the sensitivity analysis.

These quantitative results – r_{min} and $|r|_{min}$ – are completely objective, in the sense that any analyst using the same model and data will replicate precisely these numerical results. And their interpretation, so as to characterize a particular hypothesis test rejection as 'robust' versus 'fragile', can in some cases be quite clear-cut. In other cases, these objective results can call out for subjective decision making, as discussed in Section 3.5. But these sensitivity results at the very least provide an objective indication as to the situation that one is in. This inferential predicament is analogous – albeit, the reader is cautioned, not identical – to the commonly-encountered situation where the P-value at which a particular null hypothesis can be rejected is on the margin of the usual, culturally-mediated rejection criterion: This P-value objectively summarizes the weight of the evidence again the null hypothesis, but our decision as to whether or not to reject this null hypothesis is clearly to some degree subjective.

---

[32]In particular, our procedure readily extends to linear regression models with autocorrelated and/or heteroscedastic errors, and to fixed-effects panel data models.

Notably, tests with regard to some null hypothesis restriction (or set of restrictions) on the model parameters may be quite robust, whereas tests of other restrictions may be quite fragile, or fragile with respect to endogeneity in a different explanatory variable. This, too, is useful information which the sensitivity analysis proposed here can provide.

If one or more of 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, the sensitivity analysis shows that some of 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 some of the more economically important inferences, then one at least – as noted above – has an objective indication as to the predicament that one is in.

Finally, we note – per the discussion in Section 3.6 – that in some settings the analytic interest is focused on the size of a particular structural coefficient, rather than on the P-value at which some particular null hypothesis (with regard to restrictions on the components of the full vector of structural coefficients) can be rejected. In such instances what we have here called 'parameter-value-centric' sensitivity analysis is especially useful.

This 'parameter-value-centric' sensitivity analysis is a simple variation on the 'hypothesis-testing-centric' sensitivity analysis described in Sections 3.1 to 3.5 above. In Section 3.6 the sampling distribution for the OLS-estimated structural parameter vector derived in Section 2 (for a specified degree of endogeneity-covariance) is used to both obtain a confidence interval for this particular structural coefficient and also to obtain a consistent estimator for the entire vector of structural coefficients. This consistent parameter estimator makes the structural model errors asymptotically available, allowing the specified endogeneity-covariance vector

to be converted into the more-interpretable implied endogeneity-correlation vector, denoted $\hat{\rho}_{X\varepsilon}$ above.

Our 'parameter-value-centric' sensitivity analysis then simply consists of plotting this estimated confidence interval against $\hat{\rho}_{X\varepsilon}$. Such plots are easy to make and simple to interpret, so long as only one or two of the explanatory variables are allowed to be possibly endogenous; the empirical example given here (in Section 4) provides several illustrative examples of such plots. These plots are two-dimensional where endogeneity is allowed for in only one explanatory variable; they are three-dimensional where two of the explanatory variables are possibly-endogenous. In the unusual instance where a still higher-dimensional sensitivity analysis seems warranted, such confidence-interval versus $\hat{\rho}_{X\varepsilon}$ plots are impossible to graphically render and visualize, but such results could be tabulated.[33]

Whereas hypothesis test rejection P-values become increasingly uninformative as the sample length becomes large, the set of confidence intervals produced by our 'parameter-value-centric' sensitivity analysis becomes only more informative as the sample length increases.[34] We consequently conjecture that this 'parameter-value-centric' sensitivity analysis can potentially be of great value in the analysis of the impact of neglected-endogeneity in the increasingly-common empirical models employing huge data sets – e.g., Rasmussen et al. (2011) and Arellano, et al. (2018) – once our methodology is extended to the nonlinear estimation methods most commonly used in such settings.

References

[1] Arellano, M., Blundell, R. and Bonhomme, S. 2018. "Nonlinear Persistence and Partial Insurance: Income and Consumption Dynamics in the PSID", *AEA Papers and Proceedings, Volume*, 108, 281-286. https://doi.org/10.1257/pandp.20181049.

[2] 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.

---

[33]This kind of tabulation of our sensitivity analysis results was first suggested in Kiviet (2016).

[34]We also note that the 'parameter-value-centric' sensitivity analysis eliminates the specification (necessary in the 'hypothesis-testing-centric' sensitivity analysis) of any particular value for the significance level.

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

[4] Ashley, R., and Parmeter, C. F. 2019. "Sensitivity Analysis of OLS Multiple Regression Inference with Respect to Possible Linear Endogeneity in the the Explanatory Variables," working paper, available from the authors.

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

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

[7] Efron, B. 1979. "Bootstrapping Methods: Another Look at the Jackknife", *Annals of Statistics*, 7, 1-26.

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

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

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

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

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

[13] Kiviet, J. F. 2018. "Testing the Impossible: Identifying Exclusion Restrictions," *Journal of Econometrics*, forthcoming.

[14] 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.

[15] 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.

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

[17] Leamer, E. 1983. "Let's Take the Con Out of Econometrics," *American Economic Reivew*, 73, 31-43.

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

[19] 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.

[20] McCloskey, D. N. and S. T. Ziliak. 1996. "The Standard Error of Regressions," *Journal of Economic Literature XXXIV*, 97-114.

[21] Rasmussen, L. H., Larsen, T. B., Due, K. M., Tønneland, A., Overvad, K. and Lip, G. Y. H. 2011. "Impact of Vascular Disease in Predicting Stroke and Death in Patients with Atrial Fibrillation: the Danish Diet, Cancer and Health Cohort Study," *Journal of Thrombosis and Homeostasis*, 9(7), 1301-1307. https://doi.org/10.1111/j.1538-7836.2011.04308.x.