

Instrumental Variables Regressions with Uncertain Exclusion Restrictions: A Bayesian Approach

Forthcoming, Journal of Applied Econometrics

Aart Kraay
The World Bank

First Draft: May 2008

This Draft: July 2009

Abstract: The identification of structural parameters in the linear instrumental variables (IV) model is typically achieved by imposing the prior identifying assumption that the error term in the structural equation of interest is orthogonal to the instruments. Since this exclusion restriction is fundamentally untestable, there often are legitimate doubts about the extent to which the exclusion restriction holds. In this paper I illustrate the effects of such prior uncertainty about the validity of the exclusion restriction on inferences based on linear IV models. Using a Bayesian approach, I provide a mapping from prior uncertainty about the exclusion restriction into increased uncertainty about parameters of interest. Moderate prior uncertainty about exclusion restrictions can lead to a substantial loss of precision in estimates of structural parameters. This loss of precision is relatively more important in situations where IV estimates appear to be more precise, for example in larger samples or with stronger instruments. I illustrate these points using several prominent recent empirical papers that use linear IV models. An accompanying electronic table allows users to readily explore the robustness of inferences to uncertainty about the exclusion restriction in their particular applications.

1818 H Street NW, Washington DC, 20433, akraay@worldbank.org. I would like to thank Daron Acemoglu, Laura Chioda, Frank Kleibergen, Dale Poirier, Luis Servén, and two anonymous referees for helpful comments. The opinions expressed here are the author's and do not reflect the official views of the World Bank, its Executive Directors, or the countries they represent.

1. Introduction

The identification of structural parameters in the linear instrumental variables (IV) model is typically achieved by imposing the prior identifying assumption that the error term in the structural equation of interest is orthogonal to the instruments. Since this exclusion restriction is fundamentally untestable, careful empirical papers devote considerable effort to selecting clever instruments and arguing for the plausibility of the relevant exclusion restriction. Despite these efforts, authors (and readers) of these papers may in many cases legitimately entertain doubts about the extent to which the exclusion restriction holds.¹

The difficulty with standard approaches to estimation and inference in the linear IV regression model is that they do not admit any such uncertainty about the exclusion restriction but rather proceed as if it were literally true in the population of interest. In this paper I explore the implications of replacing the standard identifying assumption that the exclusion restriction is literally true with a weaker one: that there is prior uncertainty over the correlation between the instrument and the error term, captured by a well-specified prior distribution centered on zero. The standard and stark prior identifying assumption is that this distribution is degenerate with all of the probability mass concentrated at zero, so that the exclusion restriction holds with probability one in the population of interest. In most applications, however, a better approximation to the actual prior beliefs of researchers would allow for some possibility that the exclusion restriction is not exactly true, even if on average it is.

Not surprisingly, allowing for greater prior uncertainty about the exclusion restriction will reduce the precision of inferences about structural parameters of interest. In this paper I show that even modest prior uncertainty about the exclusion restriction can potentially have substantial effects on the reliability of inferences about structural parameters. This loss of precision is relatively more important in situations in which the usual IV estimator would otherwise appear to be more precise, for example, when the sample size is large or the instrument is particularly strong. The intuition for this is

¹ Murray (2006) poetically refers to this as the "cloud of uncertainty that hovers over instrumental variable estimation". See also Deaton (2009) for a recent extensive discussion of the implausibility of exclusion restrictions in many contexts.

straightforward. If there is prior uncertainty about the validity of the exclusion restriction, having a stronger instrument or a larger sample size will not reduce this uncertainty, as the data are fundamentally uninformative about its validity. Since prior uncertainty about the exclusion restriction is unaffected by sample size or the strength of the instrument, while the precision of the IV estimator improves with sample size and the strength of the instrument for the usual reasons, the effects of prior uncertainty about the exclusion restriction become relatively more important in circumstances where the IV estimator would otherwise appear to be more precise.

I illustrate these points using the Bayesian approach to inference in the context of the canonical just-identified linear IV regression model with a single endogenous regressor and a single instrument. With its explicit treatment of prior beliefs about parameters of interest, the Bayesian approach provides a natural framework for considering prior uncertainty about the exclusion restriction. In order to focus attention on the effects of varying degrees of prior uncertainty about the exclusion restriction, I will assume that prior beliefs about the other parameters of interest are appropriately diffuse. In the extreme case where there is no uncertainty at all about the exclusion restriction, this will ensure that Bayesian inferences mimic frequentist ones for the usual textbook reasons. As prior uncertainty about the exclusion restriction increases, inferences about the structural slope coefficient become less precise, and as noted, do so in a way that depends intuitively on the characteristics of the sample at hand. To aid applied researchers in quickly exploring the consequences of uncertainty about the exclusion restriction in their particular applications, I provide extensive tabulations of posterior probability intervals for alternative combinations of sample characteristics and varying degrees of uncertainty about the exclusion restriction.²

I illustrate these results using three prominent papers that use linear IV regressions. Rajan and Zingales (1998) study the relationship between financial development and growth, using measures of legal origins and institutional quality as instruments for financial development. Frankel and Romer (1999) study the effects of trade on levels of development across countries, using the geographically-determined component of trade as an instrument. Finally, Acemoglu, Johnson and Robinson (2001) study the effects of institutional quality on development in a sample of former colonies,

² A file containing these tabulations is available at <http://econ.worldbank.org/staff/akraay>.

using estimates of historical settler mortality rates in the 18th and 19th centuries as instruments. In all three cases, it is reasonable to entertain some doubts as to the literal validity of the exclusion restriction, and this will reduce the precision of inferences about the slope parameters of interest. For the first two papers I find that moderate uncertainty about the exclusion restriction is sufficient to call into question whether the estimated slopes are indeed significantly different from zero at conventional levels, while the findings of the third paper appear to be more robust to all but extreme prior uncertainty about the exclusion restriction.

This paper contributes to a small but growing literature on inference with imperfect instrumental variables. It is well-known that violations of the exclusion restriction will lead to biases in the IV estimator, and these biases will be larger the weaker are the instruments. Motivated by this observation, an early approach in the literature was to consider ranges of known deviations of the exclusion restriction, and to check the robustness of IV inferences to these deviations (see for example Angrist and Krueger (1994) and Angrist, Imbens and Rubin (1996)). This results in a range or a set of bounds on the parameters of interest.³ The limitation of this early approach is that the precision of inferences about structural parameters of interest depends not only on the magnitude of deviations from the exclusion restriction, but also how likely deviations of different magnitudes are. If for example a researcher's prior belief is that an instrument is most likely valid in the sense that large violations of the exclusion restriction are less likely than small violations, then more precise inferences will be possible than if the researcher believes that all possible violations of the exclusion restriction are equally likely.

One approach to incorporating this additional information is to assume that violations of the exclusion restriction are proportional to the sampling variation in the

³ Closely related to this approach is Nevo and Rosen (2008) who show that relatively weak assumptions on the direction and magnitude of the correlation between the instrument and error term are sufficient to provide bounds for the true value of the structural slope coefficient of interest. Also related is Small (2007) who shows how to use information in the tests of overidentifying restrictions to provide bounds for structural slope coefficients when instruments fail to satisfy the exclusion restriction. Finally, in the Angrist-Krueger dataset itself, Hoogerheide and Van Dijk (2006) apply Bayesian methods and allow for partial violations of the exclusion restriction in the sense that they allow quarter-of-birth to have a direct effect in the wage equation. They show that this increases the imprecision of inferences about the effects of education on wages in this particular context by weakening the instruments.

estimator. Restricting attention to such 'local' violations of the exclusion restriction is a useful technical device which allows a characterization of the asymptotic properties of the IV estimator with imperfect instruments, as is done by Hahn and Hausmann (2006) and Berkowitz, Caner and Fang (2008a,b). An alternative and perhaps more natural approach taken in this paper is to explicitly incorporate researchers' prior degree of confidence in the exclusion restriction into inference about parameters of interest. As noted, the Bayesian approach to inference adopted here provides a natural framework for doing so.

The results in this paper are most closely related to (although developed independently of) those in Conley, Hansen, and Rossi (2008). They study linear IV regression models in which there are potentially failures of the exclusion restriction (which they refer to as "plausible exogeneity"). They propose a number of strategies for investigating the robustness of inference in the presence of potentially invalid instruments, including a fully-Bayesian approach like the one taken here. In contrast with this paper, theirs discusses the Bayesian approach to uncertain exclusion restrictions at a quite high level of generality that may be less accessible to applied researchers interested in deploying these techniques.

While similar in approach, this paper complements Conley, Hansen and Rossi (2008) in three respects. First, I focus on the special but canonical case of a linear IV regression model with a single endogenous regressor and a single instrument. I moreover restrict attention to the case of homoskedastic disturbances that are jointly normally distributed with the instrument. While these assumptions are clearly restrictive and rule out realistic situations of heteroskedasticity and/or discrete instruments, they come with the benefit that it is possible to obtain analytic results on the effects of prior uncertainty about the exclusion restriction. This in turn helps to develop some key insights. My objective in this paper is not to develop a fully general Bayesian approach to the problem of inference with imperfect instruments. Rather my much more modest objective is to illustrate the consequences of a minimal deviation from the traditional frequentist approach to the IV model -- allowing for uncertainty about the exclusion restriction -- while otherwise restricting the Bayesian approach to mimic the traditional frequentist approach as closely as possible.

Second and closely related, given my simplifying assumptions I am able to characterize how the consequences for inference of prior uncertainty about the exclusion restriction depend on the properties of the observed sample. This also permits an extensive tabulation of the effects of prior uncertainty about the exclusion restriction for the common case of a single endogenous regressor and a single instrument that can very easily be used by applied researchers. Finally, I provide several macroeconomic cross-country applications of this approach that complement the more microeconomic examples in their paper.

2 The Linear IV Regression Model with Uncertain Exclusion Restrictions

2.1 Preliminaries

I consider the canonical linear IV regression model with a single potentially endogenous regressor and a single instrument. The structural form of the model is:

$$(1) \quad \begin{aligned} y_i &= \beta \cdot x_i + \varepsilon_i \\ x_i &= \Gamma \cdot z_i + v_i \end{aligned}$$

I assume that the regressor x and the instrument z are normalized to have zero mean and unit standard deviation, and that the error terms also have zero mean. Extensions to the case of multiple endogenous variables and multiple instruments are discussed below. My focus is on the consequences of uncertainty about the exclusion restriction for inferences about the structural slope parameter of interest, β . For the IV estimation problem to be interesting, I assume that x is endogenous, i.e. $E[x \cdot \varepsilon] \neq 0$. In contrast, I assume that $E[z \cdot v] = 0$ in the first-stage relationship.

As is well-known, the structural slope parameter β is not identified without further prior information, usually in the form of restrictions on the other parameters of the model. To see this intuitively, consider the relationship between the covariances in the data and the parameters of the model:

$$\begin{aligned}
& E[x \cdot z] = \Gamma \\
(2) \quad & E[y \cdot z] = \beta \cdot \Gamma + E[\varepsilon \cdot z] \\
& E[y \cdot x] = \beta + \Gamma \cdot E[\varepsilon \cdot z] + E[\varepsilon \cdot v]
\end{aligned}$$

where $E[\varepsilon \cdot v]$ captures the strength of the endogeneity problem that the IV estimator aims to address, and $E[\varepsilon \cdot z]$ captures the covariance between the instrument and the structural error term. Since there are four parameters on the right-hand-side of Equation (2) (β , Γ , $E[\varepsilon \cdot v]$, and $E[\varepsilon \cdot z]$) but only three observable covariances on the left-hand side, additional information is required to identify the slope coefficient of interest, β .

The standard approach is to make the prior identifying assumption that $E[\varepsilon \cdot z] = 0$, i.e. that the instrument is orthogonal to the error term in the structural equation. This is the exclusion restriction which states that the instrument has no effect on the endogenous variable other than its indirect effect through x . With this conventional identifying assumption, the structural slope coefficient can be retrieved from the ratio of the first two expressions in Equation (2). This is in fact precisely the intuition for Indirect Least Squares: the 2SLS estimator of the structural slope coefficient is the ratio of the OLS estimates of a regression of the dependent variable on the instrument (which has a probability limit $E[y \cdot x]$) to the OLS estimate of a regression of the endogenous variable on the instrument (which has a probability limit $E[x \cdot z]$).

The use of prior information to identify structural parameters in the IV regression model is thus standard practice. The point of this paper however is that the prior assumption that $E[\varepsilon \cdot z] = 0$ is likely to be a poor approximation to the actual prior beliefs of empirical researchers in many applications. As discussed in the introduction, in many cases researchers may legitimately entertain doubts as to the validity of the exclusion restriction in their particular application, even if their best guess is that it is correct. The standard prior assumption that $E[\varepsilon \cdot z] = 0$ however does not admit any such uncertainty, but rather caricatures researchers' prior beliefs as holding the exclusion restriction to be literally true in the population of interest. In contrast, researchers in many cases might assign some prior probability to values of this covariance that are different from zero. Fortunately, however, such a weaker prior

admitting uncertainty about the exclusion restriction will still be sufficient to identify the structural slope parameter of interest, but naturally it will do so less precisely. In the remainder of Sections 2 and 3 I suggest a weak prior exclusion restriction and quantify its effect on the precision of inferences regarding the structural slope parameter.

2.2 The Reduced-Form Likelihood Function

Bayesian analysis of the linear IV regression model can conveniently be based on the reduced-form of the structural model, which comes from substituting the second structural equation into the first:

$$(3) \quad \begin{aligned} y_i &= \gamma \cdot z_i + u_i \\ x_i &= \Gamma \cdot z_i + v_i \end{aligned}$$

where $u_i \equiv \varepsilon_i + \beta \cdot v_i$ and $\gamma \equiv \beta \cdot \Gamma$. The reason for this is straightforward: the reduced-form model in Equation (3) is simply a multivariate linear regression model that admits a natural conjugate prior distribution when the reduced-form disturbances are assumed to be normally distributed. This means that the prior and posterior distributions have the same form, and there are analytic results providing the mapping from the parameters of the prior distribution to the parameters of the posterior distribution, making transparent the use of the observed data to update prior beliefs regarding the reduced-form parameters.⁴ Moreover, since there is a one-to-one mapping between the structural and the reduced-form parameters, the familiar prior and posterior distributions for the reduced form parameters in the multivariate regression model in Equation (3) induce well-behaved prior and posterior distributions over the structural-form parameters as well (see Hoogerheide, Kleibergen and Van Dijk (2007)). I follow this approach, but extending it to allow for prior uncertainty over the validity of the exclusion restriction.

In order to be able to exploit these textbook results on the natural conjugate priors for the multivariate regression model, I assume that the reduced form errors and the instrument are jointly normally distributed:⁵

⁴ See for example Zellner (1971), Ch. 8 and Poirier (1996), Ch. 10.

⁵ Normality of the disturbances is a common assumption in much of the literature on Bayesian inference in the IV regression model. Here I further assume that the instrument is also normally

$$(4) \quad \begin{pmatrix} u_i \\ v_i \\ z_i \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \theta \cdot \sigma_u \cdot \sigma_v & \phi \cdot \sigma_u \\ & \sigma_v^2 & 0 \\ & & 1 \end{pmatrix} \right)$$

The key non-standard assumption here is that the correlation ϕ between the reduced-form residual u and the instrument z is non-zero. In particular, note that $\phi \cdot \sigma_u = E[u \cdot z] = E[\varepsilon \cdot z]$ will be non-zero when there are failures of the exclusion restriction.

Let Y denote the $T \times 2$ matrix with the T observations on (y_i, x_i) as rows; let Z denote the $T \times 1$ vector containing the T observations on z_i ; and recall that Z has been normalized such that $Z'Z = T$. Let $\Omega(\phi) = \begin{pmatrix} \sigma_u^2 \cdot (1 - \phi^2) & \theta \cdot \sigma_u \cdot \sigma_v \\ & \sigma_v^2 \end{pmatrix}$ denote the variance-covariance matrix of (y_i, x_i) conditional on z_i . Define the 1×2 matrix $G \equiv (\gamma : \Gamma)$ and let $\hat{G} \equiv (\hat{\gamma} : \hat{\Gamma}) = (Z'Z)^{-1} Z'Y$ denote the matrix of OLS estimates of the reduced-form slope coefficients and let $S \equiv (Y - Z\hat{G})'(Y - Z\hat{G}) / (T - 1)$ denote the estimated variance-covariance matrix of the residuals from the OLS estimation of the reduced-form slopes. With this notation, the likelihood function for the reduced-form is:⁶

$$(5) \quad L(Y, X, Z; G, \Omega(\phi), \phi) = \pi^{-TM/2} \cdot |\Omega(\phi)|^{-T/2} \cdot \exp \left[-\frac{1}{2} \text{tr} \left\{ \Omega(\phi)^{-1} \left((T-1) \cdot S + T \cdot \left(G - \left(\hat{G} - (\phi \cdot \sigma_u : 0) \right) \right)' \left(G - \left(\hat{G} - (\phi \cdot \sigma_u : 0) \right) \right) \right) \right\} \right]$$

distributed, so that failures of the exclusion restriction can be simply parameterized. More general approaches that make less restrictive distributional assumptions are also available (see for example Conley, Hansen, McCulloch and Rossi (2009)).

⁶ See for example Zellner (1973), Equation 8.6 or Poirier (1996), Equation 10.3.12 for the case where $\phi=0$.

Given the linearity of the model, the assumption of normality for the reduced form implies that the structural disturbances and the instrument are also jointly normally distributed as follows:

$$(6) \quad \begin{pmatrix} \varepsilon_i \\ v_i \\ z_i \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\varepsilon^2 & \lambda \cdot \sigma_\varepsilon \cdot \sigma_v & \rho \cdot \sigma_\varepsilon \\ & \sigma_v^2 & 0 \\ & & 1 \end{pmatrix} \right)$$

and some simple arithmetic gives the following one-to-one relationship between the parameters of the reduced-form and structural models:

$$(7) \quad \begin{aligned} \gamma &= \beta \cdot \Gamma \\ \Gamma &= \Gamma \\ \sigma_\varepsilon^2 &= \sigma_\varepsilon^2 \\ \sigma_u^2 &= \sigma_\varepsilon^2 + 2\beta\lambda\sigma_\varepsilon\sigma_v + \beta^2\sigma_v^2 \\ \theta &= \frac{\lambda\sigma_\varepsilon + \beta\sigma_v}{\sqrt{\sigma_\varepsilon^2 + 2\beta\lambda\sigma_\varepsilon\sigma_v + \beta^2\sigma_v^2}} \\ \phi &= \frac{\rho\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2 + 2\beta\lambda\sigma_\varepsilon\sigma_v + \beta^2\sigma_v^2}} \end{aligned}$$

This one-to-one mapping between the structural and reduced-form parameters plays an important role in what follows. I begin by specifying a prior distribution over the reduced-form parameters on the left-hand side of Equation (7). This mapping then implies a prior distribution over the structural parameters on the right-hand side of Equation (7). Next I can use textbook analytic results to obtain the posterior distribution for the reduced-form parameters of the model. Finally, this mapping will again imply a posterior distribution over the structural parameters of interest.

2.3 The Prior Distribution

The premise of this paper is that researchers' actual prior beliefs typically incorporate some uncertainty about the exclusion restriction, and that this uncertainty should be modelled explicitly. This in turn implies a prior belief that there is uncertainty

as to whether the reduced-form parameter ϕ is identically equal to zero or not. I approximate this prior uncertainty about the exclusion restriction with the following prior distribution over the correlation between the reduced-form error and the instrument, ϕ :

$$(8) \quad g(\phi) \propto (1 - \phi^2)^\eta$$

over the support $(-1, 1)$ where η is a parameter that governs prior confidence as to the validity of the identifying assumption. In particular, when $\eta=0$, the prior is uniformly distributed over $(-1, 1)$. As η increases the prior becomes more concentrated around zero, and in the limit we approach the standard assumption that $\phi=0$ with probability one. Figure 1 plots this prior distribution for alternative values of η . The top panel of Table 1 reports the 5th and 95th percentiles of this distribution for alternative values of η . For example, setting $\eta=500$ corresponds to the rather strong prior belief that there is a 90 percent probability that ρ is between -0.05 and 0.05 , and only a 10 percent probability that it is further away from zero.

I assume that prior beliefs for the remaining reduced-form parameters are diffuse. Although it is straightforward, and probably also realistic in many cases, to assume that researchers have non-diffuse prior beliefs about the other parameters of the model as well, the advantage of the diffuse-prior assumption is that it will deliver posterior inferences for the structural slope coefficient of interest that are very similar to frequentist ones. In particular, in the conventional but extreme case where η becomes large so that the prior for ϕ collapses to a point mass at zero, we retrieve the standard result that the posterior distribution for the reduced-form slopes will be a multivariate-t distribution centered on their OLS estimates, and with variance equal to the OLS variance estimator. In this case 95-percent confidence intervals for the reduced-form slopes can be interpreted from a Bayesian standpoint as the 2.5th and 97.5th percentiles of the posterior distribution of the reduced-form slopes.⁷

⁷ In fact, this example highlights the strong asymmetry in implicit prior beliefs that are conventionally used in the analysis of the linear IV regression model. Researchers conventionally behave as if they have completely diffuse or uninformative priors about all of the parameters of the reduced-form model, with the exception of the correlation between the instrument and the error term where the implicit prior is highly informative to the point of being degenerate at zero.

I capture these diffuse prior beliefs using a Jeffrey's prior conditional on ϕ for the remaining parameters of the structural form, i.e. $g(\mathbf{G}, \Omega(\phi) | \phi) \propto |\Omega(\phi)|^{-3/2}$. This implies the following joint prior distribution for the reduced-form parameters:

$$(9) \quad g(\mathbf{G}, \Omega(\phi), \phi) \propto |\Omega(\phi)|^{-3/2} \cdot (1 - \phi^2)^\eta$$

This joint prior distribution for the reduced-form parameters induces a prior distribution over the structural-form parameters. Of particular interest here is the implied marginal prior distribution for the correlation between the structural disturbance and the

instrument, ρ . Since $\rho = \frac{\phi \sigma_u}{\sigma_\varepsilon}$, the prior distribution of ρ will in general depend on the

entire joint prior distribution of all of the reduced-form parameters. However, since the prior distribution of the reduced-form parameters other than ϕ is chosen to be uninformative, it is straightforward to verify numerically that the marginal prior distribution of ρ has a very similar shape and percentiles as the distribution of ϕ .⁸ As a result, the percentiles of the distribution of ϕ reported in Table 1 are very similar to the percentiles of the distribution of ρ .

Finally, it is worth emphasizing that the marginal prior distribution I have chosen for ϕ (and ρ) is symmetric around zero. This assumption can easily be relaxed to allow for a prior that captures the possibility of systematic violations of the exclusion restriction, in the sense that the prior distribution for ϕ (and ρ) has a non-zero mean. However, I do not pursue this here as this extension is inessential to my main point, which is the consequences for inference of uncertainty about the exclusion restriction.

⁸ It is straightforward although tedious to compute the Jacobian of the mapping from the structural parameters to the reduced-form parameters, and use this to write down the joint prior distribution of all the structural-form parameters. It does not however appear to be tractable to extract analytically from this the implied marginal distribution of ρ . This is why I instead characterize this distribution numerically.

2.4 The Posterior Distribution

The posterior distribution for the structural form parameters is proportional to the product of the likelihood function in Equation (5) and the prior distribution in Equation (9). Multiplying these two distributions and rearranging gives:

$$(10) \quad L(\mathbf{G}, \Omega(\phi), \phi | \mathbf{Y}, \mathbf{X}, \mathbf{Z}) \propto$$

$$|\Omega(\phi)|^{-1/2} \cdot \exp \left[-\frac{1}{2} \left(\mathbf{G} - \left(\hat{\mathbf{G}} - \left(\frac{\phi \cdot \omega_{11}}{\sqrt{1-\phi^2}} : \mathbf{0} \right) \right) \right)' \left(\frac{\Omega(\phi)}{\mathbf{T}} \right)^{-1} \left(\mathbf{G} - \left(\hat{\mathbf{G}} - \left(\frac{\phi \cdot \omega_{11}}{\sqrt{1-\phi^2}} : \mathbf{0} \right) \right) \right) \right]$$

$$\cdot |\Omega(\phi)|^{-(T-2)/2} \cdot \exp \left[-\frac{1}{2} \text{tr} \left\{ \Omega(\phi)^{-1} \mathbf{S} \cdot (\mathbf{T} - 1) \right\} \right] \cdot (1 - \phi^2)^n$$

The first line is proportional to a normal distribution for the reduced-form slopes,

$\mathbf{G} \equiv (\gamma : \Gamma)$, conditional on ϕ and Ω , with mean $\left(\hat{\gamma} - \frac{\phi \cdot \omega_{11}}{\sqrt{1-\phi^2}} : \hat{\Gamma} \right)$ and variance-

covariance matrix $\frac{\Omega(\phi)}{\mathbf{T}}$. When $\phi=0$, this retrieves the standard Bayesian result for the

multivariate linear regression model with a diffuse prior for the reduced-form of the IV regression. In particular, when $\phi=0$, the posterior distribution of the reduced-form slopes conditional on the covariance matrix of the disturbances, Ω , is normal and centered on the OLS estimates of these slopes, $(\hat{\gamma} : \hat{\Gamma})$. However, when ϕ is different from zero, the mean of the conditional posterior distribution for γ needs to be adjusted to reflect this failure of the exclusion restriction, which induces a correlation between the regressor and the error term in the first structural equation. If the correlation between the regressor and the error term is positive (negative), then intuitively, the posterior mean needs to be adjusted downwards (upwards) from the OLS slope estimator. In contrast, no adjustment is required for the conditional mean of Γ , since by assumption the error term in the first-stage regression is orthogonal to the instrument.

The second line is the joint posterior distribution of $\Omega(\phi)$ and ϕ , consisting of the product of an inverted Wishart distribution for $\Omega(\phi)$ conditional on ϕ , and the posterior distribution for ϕ . The posterior inverted Wishart distribution for Ω is also a standard result, and this posterior distribution intuitively is centered on the OLS variance

$$\text{estimator, i.e. } E[\Omega(\phi)] = \frac{T-1}{T-3} \cdot S .$$

The only novel part of Equation (10) is the posterior distribution for ϕ , which is identical to the prior distribution. The prior and the posterior are identical because the data are uninformative about this unidentified parameter. However, by explicitly incorporating prior uncertainty about the exclusion restriction, it will now be possible to explicitly average over this uncertainty when making inferences about the slope coefficients of interest.

3 Inference with Uncertain Exclusion Restrictions

3.1 Inference About Reduced-Form Parameters

Some useful insights can be gained from first considering how the posterior distribution of the reduced-form slopes is affected by prior uncertainty about the exclusion restriction. When $\phi=0$, the covariance parameters in Ω can be analytically integrated out of the joint posterior distribution, resulting in a posterior multivariate-t distribution for the reduced-form slopes that is centered on their OLS estimates. This provides the textbook Bayesian interpretation of the frequentist OLS estimates and confidence intervals for the structural form. Although this convenient analytical result does not go through when $\phi \neq 0$, it is straightforward to characterize the mean and variance of the posterior distribution of the reduced-form slopes using the law of iterated expectations.

The unconditional posterior mean of the reduced-form slope coefficients is:

$$(11) \quad E[\gamma : \Gamma] = \left(\hat{\gamma} - s \cdot B(T) \cdot E \left[\frac{\phi}{\sqrt{1-\phi^2}} \right] : \hat{\Gamma} \right) = (\hat{\gamma} : \hat{\Gamma})$$

where $B(T) \equiv \frac{\Gamma((T-2)/2)}{\Gamma((T-1)/2)} \cdot \sqrt{\frac{T-1}{2}} \rightarrow 1$ as T becomes large, and the expectation on the right-hand side of Equation (11) is with respect to the marginal posterior distribution of ϕ . Consider first the posterior mean of the slope coefficient from the reduced-form regression of the endogenous variable y on the instrument z , i.e. $\gamma \equiv \beta \cdot \Gamma$. Under the conventional identifying assumption that ϕ is identically equal to zero, the instrument z is uncorrelated with the error term in this regression. We then have the usual Bayesian diffuse-prior result that the mean of the posterior distribution of γ is its OLS slope estimate, $\hat{\gamma}$. However, when there is prior (and thus also posterior) uncertainty about ϕ , there is an additional term in the posterior mean reflecting this uncertainty. This term involves the expectation (with respect to the posterior density for ϕ) of $\frac{\phi}{\sqrt{1-\phi^2}}$. Since we have assumed that the prior (and posterior) are symmetric around $\phi=0$, this term is unsurprisingly zero in expectation. If we are agnostic as to whether the correlation between the instrument and the error term is positive or negative, on average this does not affect the posterior mean of γ . Of course for other priors (and posteriors) not symmetric around zero this would not be the case, and the posterior mean of γ would have to be adjusted accordingly. Finally observe that of course prior uncertainty about the exclusion restriction does not matter for estimates of the first-stage slope coefficient Γ .

The effects of uncertainty about the exclusion restriction on the posterior variance are substantively more interesting. The posterior unconditional variance of the reduced-form slopes is:

$$(12) \quad V[\gamma : \Gamma] = \begin{pmatrix} s_{11}(1/T + E[\phi^2/(1-\phi^2)]) & s_{12}/T \\ s_{12}/T & s_{22}/T \end{pmatrix} \cdot \begin{pmatrix} T-1 \\ T-3 \end{pmatrix}$$

Prior uncertainty about the exclusion restriction is reflected in an increase in the posterior variance of the reduced-form slope coefficient, γ . This variance consists of the

usual component that declines with sample size, s_{11}/T , as well as an adjustment capturing uncertainty about the exclusion restriction, $s_{11} \cdot E[\phi^2/(1-\phi^2)]$. This second term is recognizable as the variance of the adjustment to the posterior mean noted above. A key observation here is that this adjustment does not decline with sample size, and so uncertainty about the exclusion restriction has proportionately larger effects on the posterior variance of the reduced-form slope coefficient γ when the sample size is large. In contrast, there is no change in the posterior variance of the slope coefficient from the first-stage regression, Γ , as potential violations of the exclusion restriction are not relevant to the estimation of this slope parameter.

This increase in the posterior variance of the reduced-form slope γ is potentially quite important quantitatively as it does not decline with sample size. The bottom panel of Table 1 illustrates the point. Define $\left(1 + T \cdot E\left[\frac{\phi^2}{1-\phi^2}\right]\right)^{1/2}$ as the ratio of the standard deviation of the posterior distribution of γ in the case where there is prior uncertainty about ϕ , to the same standard deviation in the standard case where ϕ is identically equal to zero. This ratio captures the inflation of the posterior standard deviation due to uncertainty about ϕ . This ratio can be large, particularly in cases where the sample size is large and/or when there is greater prior uncertainty about ϕ . For example, for the case where $\eta=100$, so that 90 percent of the prior probability mass for ϕ lies between -0.12 and 0.12 , the posterior standard deviation is 22 percent higher in a sample size of 100, but 87 percent higher when the sample size is 500, and 245 percent larger in a sample of size 1000. Moving to the left in the table to cases with greater prior uncertainty about ϕ results in even greater inflation of the posterior standard deviation of the reduced-form slope.

3.2 Inference About Structural Parameters

We have seen that uncertainty about the exclusion restriction can have potentially large effects on the posterior variance of the reduced-form slope coefficient γ . This increase in posterior uncertainty will carry over to inferences about the structural slope parameter β since $\beta=\gamma/\Gamma$. While it does not appear possible to analytically

characterize the effects of this additional uncertainty on the posterior marginal distribution of β , it is straightforward to do so by numerical sampling from this distribution.⁹ A complication however is that the effects of prior uncertainty about the exclusion restriction on the posterior distribution of the structural slope coefficient of interest will be sample-dependent. This is because the posterior distribution in Equation (10) depends on the observed sample through the OLS estimates of the reduced form slopes ($\hat{\gamma}$ and $\hat{\Gamma}$) and the residual variances (S).

In order to provide intuitions for how the effects of prior uncertainty about the exclusion restriction will vary in different observed samples, I present some simple illustrative calculations for alternative hypothetical observed samples. I begin by innocuously assuming that the observed data on y are scaled to have mean zero and variance one, as are x and z . The observed sample can therefore be characterized by three sample correlations, R_{yx} , R_{yz} , and R_{xz} , and the observed reduced-form slopes and residual variances can be expressed in terms of these correlations as:

$$(13) \quad (\hat{\gamma}; \hat{\Gamma}) = (R_{yz}; R_{xz}) \quad \text{and} \quad S = \begin{pmatrix} 1 - R_{yz}^2 & R_{xy} - R_{zy} \cdot R_{xz} \\ & 1 - R_{xz}^2 \end{pmatrix}$$

I next take a hypothetical dataset summarized by a combination of assumptions on the three sample correlations and number of observations T , and then I sample from the posterior distribution of β , for a range of values for the parameter governing prior uncertainty about the exclusion restriction, η . I take 10,000 draws from the posterior distribution of β in each case, and compute the 2.5th and 97.5th percentiles of this sample of draws. This is analogous to a standard frequentist 95 percent confidence interval for the IV estimate of the slope coefficient.¹⁰

⁹ This can be done using a very simple four-step procedure: 1) draw ϕ from its marginal posterior distribution; 2) draw $\Omega(\phi)$ from its posterior Wishart distribution conditional on ϕ ; 3) draw γ and Γ from their posterior bivariate normal distribution conditional on Ω and ϕ ; and 4) use Equation (7) to retrieve the structural-form coefficients implied by the drawn reduced-form coefficients.

¹⁰ In fact, since the posterior distribution of β conditional on ϕ and Ω is a Cauchy-like ratio of correlated normal random variables, the moments of the unconditional posterior distribution of β do not exist (see Hoogerheide, Kleibergen and Van Dijk (2007)), and so I compute only the percentiles of this posterior distribution.

The results of this exercise are summarized in Table 2. Each row of the table corresponds to a set of assumptions on the observed sample correlations and the sample size. These assumptions are spelled out in the left-most columns, in italics. In each row I also report the 2.5th and 97.5th percentiles of the posterior distribution for β in the standard case where there is no uncertainty about the exclusion restriction, i.e. when $\rho=\phi=0$. This serves as a benchmark. The right-most columns correspond to various assumptions about η , capturing varying degrees of prior certainty about the exclusion restriction. I consider the same range of values as in Table 1, and for reference at the top of the table I report the 5th and 95th percentiles of the prior distribution of ϕ (and ρ) that these imply. Each cell entry reports the length of the interval from the 2.5th to the 97.5th percentile of the posterior distribution of β , expressed as a ratio to the length of this same interval when $\rho=\phi=0$, i.e. relative to the standard case with no uncertainty about the exclusion restriction.

Not surprisingly, all of the entries in Table 2 are greater than one, reflecting the fact that prior uncertainty about the exclusion restriction increases the dispersion of the posterior distribution of β . This increase in posterior uncertainty regarding β is of course higher the greater is prior uncertainty regarding the exclusion restriction. Consider for example when all three sample correlations are equal to 0.5 and the sample size is equal to 100. When $\eta=10$, corresponding to significant uncertainty about the exclusion restriction, the 95 percent confidence interval for β is 2.14 times larger than the benchmark case where $\rho=\phi=0$ by assumption. However, as η increases and there is less prior uncertainty about the exclusion restriction, this magnification of posterior uncertainty is smaller, and when $\eta=500$ the confidence intervals are just 1.03 times larger than the benchmark case.

Unsurprisingly, Table 2 also confirms that in all cases the magnification of posterior uncertainty is greater the larger is the sample size. For example, when all three sample correlations are equal to 0.5 and the sample size is equal to 100, the confidence interval for β is inflated by a factor of 2.14 when $T=100$, but it is inflated by a factor of 4.45 when $T=500$. The reason for this is straightforward: the correction to the posterior variance of the reduced-form parameter γ to account for uncertainty about the exclusion restriction does not decline with sample size, and so its effect on posterior

uncertainty regarding γ is proportionately greater the larger is the sample size. Since $\beta = \gamma/\Gamma$ this is also reflected in proportionately greater posterior uncertainty about the structural slope.

The more interesting insight from Table 2 is that the magnification of posterior uncertainty about β also depends on the moments of the observed sample in a very intuitive way. Consider the first panel of Table 2, where I vary the strength of the first-stage sample correlation between the instrument and the endogenous variable, R_{xz} , holding constant the other two correlations.¹¹ In the standard case where $\rho = \phi = 0$ by assumption, the confidence intervals of course shrink as the strength of the first-stage relationship increases. However, the magnification of the posterior variance *increases* as the strength of the first-stage relationship increases. The intuition for this is analogous to the intuition for the effects of sample size. A larger sample size, and also a stronger first-stage relationship between the instrument and the endogenous variable permit more precise inferences about β . However, a larger sample size and a stronger first-stage regression cannot reduce our intrinsic uncertainty about the validity of the exclusion restriction, and so the adjustment to the posterior variance to account for this is proportionately greater. Of course this does not mean that uncertainty about the exclusion restriction is less important in an absolute sense in small samples or with weak instruments -- only that its effects on posterior uncertainty are smaller *relative to* other sources of posterior imprecision about the parameters of interest.

The same insight holds in the second and third panels of Table 2, where I vary the strength of the observed sample correlation between the dependent variable and the instrument, R_{yz} , and the strength of the observed correlation between y and x , R_{xy} . To see why, remember that the estimated variance of the conventional IV estimator in this

¹¹ In these examples I have chosen hypothetical samples in which we are unlikely to encounter well-known weak-instrument pathologies. In fact, the minimum correlation of 0.3 between the endogenous variable and the instrument in this table is deliberately chosen to ensure that the first-stage F-statistic is almost 10 in the smallest sample of size $T=100$ that I consider, and is greater than 10 in all other cases. This corresponds to the rule of thumb proposed by Staiger and Stock (1997) for distinguishing between weak and strong instruments. These weak-instrument pathologies pose no particular difficulties for Bayesian analysis that bases inference on the entire posterior distribution of β . However, with weak instruments the Bayesian highest posterior density intervals would no longer necessarily be symmetric around the mode of the posterior distribution.

normalized and just-identified example is $T^{-1} \cdot \hat{\sigma}_\varepsilon^2 \cdot R_{xz}^{-2}$ where

$$\hat{\sigma}_\varepsilon^2 = 1 - 2 \cdot \frac{R_{xy} \cdot R_{yz}}{R_{xz}} + \left(\frac{R_{yz}}{R_{xz}} \right)^2.$$

Absent concerns about the validity of the exclusion

restriction, the IV estimator is more precise the greater is the observed sample correlation between y and x, and the smaller is the observed sample correlation between z and y. Intuitively, a high value of R_{xy} (holding constant the other two sample correlations) corresponds to a stronger structural relationship between y and x, while a low value of R_{yz} (again holding constant the other two correlations) corresponds to smaller endogeneity problems. However, since the exclusion restriction is fundamentally untestable, varying either of these correlations has no effect on the intrinsic uncertainty about the exclusion restriction, which becomes relatively more important as other sources of imprecision in the IV estimator decline. In particular, in the middle panel of Table 2 we see that magnification of the posterior probability interval declines as the correlation R_{yz} increases (i.e. as the conventional IV estimator becomes *less* precise). Similarly, in the bottom panel of Table 2 we see that magnification of the posterior probability interval increases as the correlation R_{xy} increases (i.e. as the conventional IV estimator becomes *more* precise).

For reasons of space, Table 2 reports the effects of uncertainty about the exclusion restriction only for a few illustrative hypothetical sample correlations. As a complement to this, I provide an extensive tabulation of the posterior distribution of β electronically at <http://econ.worldbank.org/staff/akraay>. There I report the 2.5th, 50th, and 97.5th percentiles of the posterior distribution of β for varying degrees of uncertainty about the exclusion restriction, for sample sizes of $T=100, 200, 500,$ and $1000,$ and for all possible sample correlations $R_{xy}, R_{yz},$ and R_{xz} (in increments of 0.1 that are consistent with a positive definite sample covariance matrix for the data). This tabulation is intended as a convenient tool for applied researchers who would like to quickly assess the robustness of their findings to uncertainty about the exclusion restriction. By choosing the appropriate sample size and sample correlations corresponding to their particular application, users can readily document the effects of uncertainty about the exclusion restriction without themselves having to resort to numerical methods.¹²

¹² The reader might wonder whether the 10,000 draws used to generate Table 2 and the accompanying more extensive tabulations are “enough” in the sense that they will produce

In summary, we have seen that prior uncertainty about the exclusion restriction can substantially increase posterior uncertainty about the key structural slope coefficient of interest, β . The magnitude of this inflation of posterior uncertainty depends of course on the degree of prior uncertainty about the exclusion restriction. But it also depends on the characteristics of the observed sample in a very intuitive way. Holding other things constant, a larger sample size, a stronger first-stage relationship between the instrument and the endogenous variable, a stronger structural correlation between dependent variable and the endogenous variable, and a weaker reduced-form correlation between the endogenous variable and the instrument all imply a more precise IV estimator, absent any prior uncertainty about the exclusion restriction. However, since none of these factors help to reduce prior (or posterior) uncertainty about the exclusion restriction, this uncertainty becomes relatively more important in such cases.

4. Empirical Applications

I next demonstrate the quantitative importance of prior uncertainty about exclusion restrictions for inference in three well-known empirical studies that use linear instrumental variables models. Acemoglu, Johnson and Robinson (2001, hereafter AJR) study the causal effects of institutions on economic development. Using a sample of 64 former colonies, they regress the logarithm of GDP per capita on a measure of property rights protection. They propose using historical data on mortality rates experienced by settlers during the colonial period as a novel instrument for institutional quality. AJR argue that in areas where settlers experienced high mortality rates, colonial powers had

reliable estimates of the percentiles of the posterior distribution of interest. One way to assess this is to successively draw samples of 10,000 using different random seeds for each sample. Doing so comprehensively for all of the entries in Table 2 and the accompanying more extensive tabulation would be computationally costly. Instead, I focus on the middle row of the each panel of Table 2, and draw 100 samples of 10,000 for the case $R_{xy}=R_{xz}=R_{yz}=0.5$ and $T=100$. I then calculate the coefficient of variation across these 100 samples of the magnification of the posterior probability intervals reported in the columns of Table 2. I find that this coefficient of variation is approximately 0.017 (averaging across the different values of η in the columns of Table 2), suggesting that 10,000 draws provides reasonably reliable results in the following sense: over repeated samples, there is approximately a 70 percent chance that the simulated magnification of the posterior probability interval is within 1.7 percent of that reported in the Table 2. This degree of precision is likely to be adequate for the intended purpose of this table, which is to provide a quick and easy check on the robustness of particular results. If more accurate results are needed, it is also straightforward to implement the sampling algorithm described in Footnote 9 for a much larger set of draws.

few incentives to set up institutions that protect property rights and provide a foundation for subsequent economic activity. In a simple bivariate specification there are a number of obvious concerns regarding the validity of the exclusion restriction that settler mortality rates matter for development only through their effects on institutional quality. Historical settler mortality rates might be correlated with the tropical location and intrinsic disease burden of a country, and these factors may matter directly for modern development. AJR seek to address such concerns in their paper through the addition of various control variables to capture these effects. For example, we will show results using one of their core specifications in which they control for latitude to capture such locational effects (Table 4, Column 2 in AJR). And in the paper they also present a wide range of results with direct controls for location and the disease burden.¹³

Nevertheless, readers of AJR might reasonably entertain some doubts as to whether the exclusion restriction holds exactly even in these extended specifications. There are many potential correlates of settler mortality rates that might in turn be correlated with development outcomes. For example, Glaeser et. al. (2004) argue that low settler mortality rates may have operated through investments in human capital rather than institutions to protect property rights. Here I do not take any stand as to which of these potential failures of the exclusion restriction is the right one. Rather I simply argue that reasonable people might entertain some uncertainty about the validity of the exclusion restriction.

My second example is Frankel and Romer (1999, hereafter FR), who study the relationship between trade openness and development in a large cross-section of countries. They regress log GDP per capita on trade as a share of GDP. To address concerns about potential reverse causation and omitted variables, they propose a novel instrument based on the geographical determinants of bilateral trade. In particular, they

¹³ Ideally I would like to use one of AJR's specifications with a more complete set of control variables to illustrate the effects of uncertainty about exclusion restrictions. However, in many of their specifications with more control variables, their instruments are much weaker, and I do not want to conflate my point about uncertainty regarding exclusion restrictions with the well-known concerns with weak instruments. For example, in Columns (7) and (8) of Table 4, AJR introduce continent dummies, and continent dummies together with latitude. In these specifications, I find first-stage F-statistics on the excluded instrument of 6.83 and 3.97, well below the Staiger and Stock (1997) rule of thumb of 10. This suggests that the settler mortality instrument does not have sufficiently strong explanatory power within geographic regions. See Albouy (2008) for a more extensive discussion of weak instruments problems in AJR.

estimate a regression of bilateral trade between country pairs on the distance between the countries in the pair, their size measured by log population and log area, and a dummy variable indicating whether either country in the pair is landlocked. They then use the fitted values from this bilateral trade regression to come up with a constructed trade share for each country that reflects only these geographical determinants of trade, which they use as an instrument for trade. In their core specification, they also control directly for country size, as measured by log population and log land area, to control for the fact that large countries tend to trade less and these size variables also enter in the bilateral trade equation. There are however various reasons why the necessary exclusion restriction (that the geographically-determined component of trade matters for development only through its effects on overall trade) may not hold exactly, that are discussed in detail in Rodríguez and Rodrik (2000).

My third example comes from Rajan and Zingales (1998, hereafter RZ), who study the relationship between financial development and growth. In contrast with the previous two papers that exploit purely cross-country variation, this paper uses a novel identification strategy that exploits within-country cross-industry differences in manufacturing growth rates. They construct a measure of the dependence of different manufacturing sectors on financial services, and then ask whether industries that are more financially-dependent grow faster in countries where financial development is greater. In particular, they estimate regressions of the growth rate of industry i in country j on a set of country dummies, a set of industry dummies, the initial size of the industry, and an interaction of the financial dependence of the sector with the level of financial development in the country. In a number of specifications, RZ instrument for this final interaction term with variables capturing the legal origins of the country and a measure of institutional quality, all interacted with a measure of financial development. In particular, I will focus on the specification in Table 4, column 6 of RZ, where the relevant measure of financial dependence is an index of accounting standards recording the types of information provided in annual reports of publicly-traded corporations in a cross-section of countries.

This third example differs from the previous ones in two key respects. First, because RZ rely on the within-country variation in sectoral growth rates, potential violations of the exclusion restriction are less obvious than in the previous two cases. In

RZ, the requirement is that the instruments be orthogonal to the country- and industry-specific component of growth, since the regressions contain country and industry dummies. Thus for example, concerns about the exclusion restriction are not that countries with faster growth adopt better accounting standards, but rather that countries with a relatively faster growth in financially-dependent industries would adopt better accounting standards. Nevertheless there might be residual concerns about the validity of the exclusion restriction even in this case. The second difference is that RZ use multiple instruments, while the results I show above apply to the case of a single instrument. To make the RZ results fit into the framework of this paper, I choose just one of their instruments and first reproduce the RZ results in this just-identified case. For this purpose I choose their index of efficiency and integrity of the legal system, produced by a commercial risk rating agency, as the one instrument of choice. Doing so gives a result that is of comparable significance to the RZ core result, although the magnitude of the estimated coefficient becomes somewhat larger than what RZ report.¹⁴

I use datasets provided by the authors to reproduce their results. In each of the three examples, I first project the dependent variable, the regressor of interest, and the instrument on all the remaining control variables that these authors treat as exogenous, so that I can identify these residuals as y , x , and z in the theoretical discussion above. I also normalize the variance of z to be equal to one, consistent with the discussion above. I then take 10,000 draws from the posterior distribution of β , for alternative values of η corresponding to varying degrees of prior uncertainty about the exclusion restriction. I then compute the 2.5th, 50th and 97.5th percentiles of this distribution.

Table 3 summarizes the results, with three panels corresponding to the three examples. In each panel in the first column I report the sample size and my replication of the relevant IV slope coefficient and standard error from each paper. In the columns of the table I provide summary statistics on the posterior distribution for the slope coefficient, for varying degrees of prior uncertainty about the exclusion restriction. In addition, Figure 2 plots the posterior densities for the slope coefficient for selected values of η . Unsurprisingly, in all three panels of this figure we clearly see how the

¹⁴ An alternative is to use just their dummy variable for Scandinavian legal origins as an instrument, which generates results that are quite similar to those reported by RZ. Conversely, using either dummies for British or French legal origins alone as an instrument does not deliver significant IV estimates of the coefficient on the interaction variable of interest.

posterior distribution of the slope coefficient becomes more dispersed as uncertainty about the exclusion restriction increases.

This increase in posterior dispersion is quantified in Table 3, which reports the 2.5th, 50th, and 97.5th percentiles of the posterior distribution of the structural slope coefficient for each of the three papers. To read this table, it is useful to begin with the last column which reports these percentiles for the limiting case where η tends to infinity and thus the prior distribution imposes $\phi=0$ with certainty. This corresponds to the standard Bayesian IV estimates in which there is no uncertainty regarding the exclusion restriction. Because of my choice of diffuse priors for all of the parameters other than ϕ , when $\phi=0$ these Bayesian results mimic the classical ones quite closely, with these percentiles quite similar to the 95 percent confidence intervals reported in the first column. This is particularly so for RZ, while for FR and AJR the posterior distribution of the slope has a somewhat longer right tail, with the result that the 97.5th percentiles are a bit higher than the upper bounds of the classical confidence intervals. This is also apparent in Figure 2, where the thin solid line plots a normal distribution with mean and standard deviation corresponding to the classical IV slope coefficient estimate and estimated standard error. For RZ this normal distribution coincides almost perfectly with the posterior distribution for the slope when $\phi=0$, while there are some small discrepancies for the other two papers.

Moving from right to left in Table 3 illustrates the effects of greater prior uncertainty about the exclusion restriction. In each of the three panels, I summarize this increase in the dispersion of the posterior distribution by reporting the length of the interval from the 2.5th percentile to the 97.5th percentile, relative to the length of the same interval when $\phi=0$ with certainty. These intervals expand substantially as uncertainty about the exclusion restriction increases. For example, for FR in the middle panel, this interval is 2.8 times as wide when $\eta=10$, while for RZ in the bottom panel it is 7.26 times as wide. This greater proportional effect on posterior uncertainty about the structural slope is consistent with what we saw in the artificial samples in Table 2, as RZ have a larger sample size and a stronger instrument than do FR. In contrast, for AJR with their smaller sample, the increase in posterior dispersion is smaller.

Table 2 also can be used to determine how great prior uncertainty about the exclusion restriction needs to be in order for the interval from the 2.5th percentile to the 97.5th percentile of the posterior distribution of β to include zero. In the case of AJR, their particular specification that we report is most robust to uncertainty about the exclusion restriction. Even when $\eta=5$, so that there is a great deal of prior uncertainty, with 90 percent of the prior probability mass for ϕ (and ρ) between -0.46 and 0.46, the 2.5th percentile of the posterior distribution of the slope is greater than zero. This is not however the case for FR and RZ. Moving from $\eta=200$ to $\eta=100$, the 2.5th percentile of the posterior distribution of the slope falls below zero. This in turn means that if the prior distribution of ϕ (and ρ) is such that more than 10 percent of the prior probability mass falls outside the interval of about (-0.1,0.1), then the Bayesian analog of the 95 percent confidence interval includes zero.

5. Conclusions and Extensions

The validity of the IV estimator depends crucially on the validity of fundamentally untestable exclusion restrictions. Identification of structural parameters of interest usually is achieved by imposing the prior identifying assumption that the exclusion restriction holds exactly. Yet in many cases a better characterization of researchers' prior beliefs is that the exclusion restriction is not exactly true in the relevant population, but only approximately so. In this paper I suggest how to explicitly incorporate prior uncertainty about the exclusion restriction into the linear IV regression model. This prior uncertainty about the exclusion restriction leads to greater posterior uncertainty about parameters of interest, in some cases quite substantially so. This enables straightforward checks of the robustness of inferences about structural parameters to varying degrees of prior uncertainty about the exclusion restriction.

There are at least two natural extensions of the results presented here. The first I have already discussed: allowing the prior distribution for the correlation between the instrument and the error term to have a non-zero mean. This would encompass not only prior uncertainty about the validity of the exclusion restriction, but also prior beliefs about the direction of likely violations of the exclusion restriction. For example, one might specify a prior distribution for ϕ that is a translation of a beta distribution, i.e.

$(\phi + 1)/2 \sim \text{Beta}(\eta_1, \eta_2)$. With appropriate choices of the prior parameters η_1 and η_2 , a prior such as this can capture prior beliefs regarding both the mean and the variance of ϕ . Since there is no updating of the prior distribution of ϕ , we will have the same posterior distribution, and we can simply (numerically) integrate over this distribution to arrive at the marginal posterior distribution for the slope coefficients of interest. This will have predictable effects on the results presented here: the posterior mean of the distribution of the structural slope coefficients will need to be adjusted to reflect the non-zero prior and posterior mean for the distribution of ϕ , since the expectation in Equation (11) will no longer be zero. This extension may be practically useful in many situations where there might be obvious potential directions for violations of the exclusion restriction.

The second is to consider the case of multiple instruments and multiple endogenous variables. In this paper, I have focused on the case of a single endogenous variable and a single instrument in order to keep the results as transparent as possible. Moving to the case of multiple endogenous variables and potential overidentification poses two further challenges. First, when there are multiple instruments, we need to elicit a multivariate prior distribution over the correlation between each of the instruments and the structural error term, rather than just a simple univariate prior over a single parameter that I have used here. In practice, it may be difficult to flexibly specify such a prior in a way that captures differing and potentially correlated degrees of certainty about the exclusion restriction for each instrument. Second, in the case of overidentification, the mapping from the reduced-form parameters to the structural parameters is more complex. In the just-identified case, the mapping between the reduced-form and structural form slopes, $\gamma = \Gamma\beta$, can immediately be inverted to retrieve the structural slopes $\beta = \Gamma^{-1}\gamma$ since the matrix of first-stage slope coefficients is square. In the over-identified case, a simple possibility is to instead use a generalized inverse to retrieve an approximation to the structural slopes, i.e. $\tilde{\beta} = (\Gamma'\Gamma)^{-1}\Gamma'\gamma$, although the properties of such an approximation are not known. For more elaborate approaches to simulating the posterior distribution of the structural slopes that is implied by the prior and posterior distribution over the reduced-form parameters in the overidentified case, see Hoogerheide, Kleibergen and Van Dijk (2007).

References

- Acemoglu, Daron, Simon Johnson, and James A. Robinson (2001). "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review*. 91(5):1369-1401.
- Albouy, David (2008). "The Colonial Origins of Comparative Development: An Investigation of the Settler Mortality Data". NBER Working Paper No. 14130.
- Berkowitz, Daniel, Mehmet Caner and Ying Fang (2008a). "Are 'Nearly Exogenous' Instruments Reliable?". *Economics Letters*. (article in press).
- Berkowitz, Daniel, Mehmet Caner and Ying Fang (2008b). "The Validity of Instruments Revisited". Manuscript, University of Pittsburgh.
- Conley, Tim, Christian Hansen and Peter E. Rossi (2008). "Plausibly Exogenous". Manuscript. Graduate School of Business, University of Chicago.
- Conley, Timothy, Christian Hansen, R.E. McCulloch, and Peter E. Rossi (2008). "A Non-Parametric Bayesian Approach to the Instrumental Variable Problem". *Journal of Econometrics*, forthcoming.
- Deaton, Angus (2009). "Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development." Keynes Lecture to the British Academy.
- Frankel, Jeffrey A. and David Romer (1999). "Does Trade Cause Growth?" *The American Economic Review*, (June) 379-399.
- Glaeser, Edward, Rafael Laporta, Florencio Lopez-de-Silanes, and Andrei Shleifer (2004). "Do Institutions Cause Growth?". *Journal of Economic Growth*. 9(3):271-303.
- Hahn, Jinyong and Jerry Hausman (2006). "IV Estimation with Valid and Invalid Instruments". *Annales d'Economie et Statistique*.
- Hoogerheide, Lennart and Herman K. van Dijk (2006). "A reconsideration of the Angrist-Krueger analysis on returns to education". Report 2006-15 of the Econometric Institute (Erasmus University Rotterdam).
- Hoogerheide, Lennart, Frank Kleibergen and Herman van Dijk (2007). "Natural Conjugate Priors for the Instrumental Variables Regression Model Applied to the Angrist-Krueger Data". *Journal of Econometrics*. 138(1):63-103.
- Murray, Michael (2006). "Avoiding Invalid Instruments and Coping with Weak Instruments". *Journal of Economic Perspectives*. 20(4):111-132.
- Nevo, Aviv and Adam Rosen (2008). "Identification with Imperfect Instruments". CEMMAP Working Paper CWP16/06.

Poirier, Dale J. (1998). "Revising Beliefs in Nonidentified Models". *Econometric Theory*. 14:483-509.

Rajan, Raghuram and Luigi Zingales (1998). "Financial Dependence and Growth". *American Economic Review*. 88(3):559-586.

Rodriguez, Francisco and Dani Rodrik (2001). "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-Country Evidence". *NBER Macroeconomics Annual*. 15:261-325.

Small, Dylan (2007). "Sensitivity Analysis for Instrumental Variables Regression With Overidentifying Restrictions". *Journal of the American Statistical Association*. 102(479):1049-1058.

Staiger, D. and J.H. Stock (1997). "Instrumental Variables Regression With Weak Instruments". *Econometrica*. 65:557-586.

Table 1: Percentiles of Prior Distribution for $\phi = \text{CORR}(z, u)$

	Value of Prior Parameter η					
	<u>5</u>	<u>10</u>	<u>100</u>	<u>200</u>	<u>500</u>	<u>1000</u>
<i>90% Prior Probability of ϕ Between:</i>						
Lower	-0.46	-0.34	-0.12	-0.08	-0.05	-0.04
Upper	0.46	0.34	0.12	0.08	0.05	0.04
<i>Inflation of Posterior Standard Deviation of γ When T=:</i>						
100	3.32	2.45	1.22	1.12	1.05	1.02
200	4.58	3.32	1.41	1.22	1.10	1.05
500	7.14	5.10	1.87	1.50	1.22	1.12
1000	10.05	7.14	2.45	1.87	1.41	1.22

Table 2: Inference in the IV Case

		Value of Prior Parameter η						
		5	10	100	200	500		
90 percent of prior probability between:								
	Lower	-0.46	-0.34	-0.12	-0.08	-0.05		
	Upper	0.46	0.34	0.12	0.08	0.05		
Assumptions on Observed Sample		Width of 95% Confidence Interval for Indicated Value of η (Relative to Width when $\phi=0$)						
Vary Strength of First-Stage CORR(x,z)								
<i>Rxy= 0.5</i>	<i>Ryz= 0.5</i>	<i>Rxz= 0.3</i>						
95% CI for β	(1.00, 4.02)		<i>T=100</i>	1.85	1.49	1.05	1.05	1.04
95% CI for β	(1.32, 2.20)		<i>T=500</i>	4.44	3.16	1.41	1.26	1.11
<i>Rxy= 0.5</i>	<i>Ryz= 0.5</i>	<i>Rxz= 0.5</i>						
95% CI for β	(0.65, 1.52)		<i>T=100</i>	2.95	2.14	1.17	1.10	1.03
95% CI for β	(0.84, 1.19)		<i>T=500</i>	6.42	4.45	1.72	1.42	1.17
<i>Rxy= 0.5</i>	<i>Ryz= 0.5</i>	<i>Rxz= 0.7</i>						
95% CI for β	(0.47, 0.99)		<i>T=100</i>	3.28	2.41	1.22	1.13	1.05
95% CI for β	(0.61, 0.83)		<i>T=500</i>	7.19	5.00	1.86	1.48	1.23
Vary Strength of Reduced Form CORR(y,z)								
<i>Rxy= 0.5</i>	<i>Ryz= 0.3</i>	<i>Rxz= 0.5</i>						
95% CI for β	(0.24, 0.99)		<i>T=100</i>	3.71	2.64	1.24	1.11	1.05
95% CI for β	(0.45, 0.76)		<i>T=500</i>	8.09	5.75	2.01	1.60	1.29
<i>Rxy= 0.5</i>	<i>Ryz= 0.5</i>	<i>Rxz= 0.5</i>						
95% CI for β	(0.65, 1.52)		<i>T=100</i>	2.95	2.14	1.17	1.10	1.03
95% CI for β	(0.84, 1.19)		<i>T=500</i>	6.42	4.45	1.72	1.42	1.17
<i>Rxy= 0.5</i>	<i>Ryz= 0.7</i>	<i>Rxz= 0.5</i>						
95% CI for β	(1.01, 2.11)		<i>T=100</i>	2.02	1.57	1.07	1.06	1.05
95% CI for β	(1.21, 1.65)		<i>T=500</i>	4.17	3.13	1.33	1.21	1.09
Vary Strength of Structural CORR(y,x)								
<i>Rxy= 0.3</i>	<i>Ryz= 0.5</i>	<i>Rxz= 0.5</i>						
95% CI for β	(0.60, 1.63)		<i>T=100</i>	2.57	1.92	1.12	1.06	1.02
95% CI for β	(0.81, 1.23)		<i>T=500</i>	5.34	3.70	1.53	1.31	1.13
<i>Rxy= 0.5</i>	<i>Ryz= 0.5</i>	<i>Rxz= 0.5</i>						
95% CI for β	(0.65, 1.52)		<i>T=100</i>	2.95	2.14	1.17	1.10	1.03
95% CI for β	(0.84, 1.19)		<i>T=500</i>	6.42	4.45	1.72	1.42	1.17
<i>Rxy= 0.7</i>	<i>Ryz= 0.5</i>	<i>Rxz= 0.5</i>						
95% CI for β	(0.72, 1.39)		<i>T=100</i>	3.65	2.70	1.28	1.16	1.07
95% CI for β	(0.87, 1.15)		<i>T=500</i>	8.14	5.73	2.03	1.64	1.28

Table 3: Empirical Examples

	Value of Prior Parameter η					
	5	10	100	200	500	∞
<i>90% Prior Probability of ρ Between:</i>						
Lower	-0.46	-0.34	-0.12	-0.08	-0.05	0.00
Upper	0.46	0.34	0.12	0.08	0.05	0.00
Acemoglu-Johnson-Robinson (2001)						
(Table 4, Column 2)						
T=64						
IV Slope = 0.96						
IV Standard Error = 0.21						
95% C.I. = (0.53, 1.39)						
<i>Posterior Distribution for Slope</i>						
2.5th Percentile	0.08	0.32	0.61	0.63	0.63	0.65
Mode	0.95	0.96	0.96	0.96	0.96	0.96
97.5th Percentile	2.31	2.06	1.80	1.81	1.76	1.75
Increase in P025-P975 range	2.02	1.57	1.08	1.07	1.02	1.00
Frankel-Romer (1999)						
(Table 3, Column 2)						
T=150						
IV Slope = 1.97						
IV Standard Error = 0.91						
95% C.I. = (0.18, 3.76)						
<i>Posterior Distribution for Slope</i>						
2.5th Percentile	-5.62	-3.61	-0.28	0.01	0.14	0.31
Mode	1.98	1.95	1.97	1.98	1.96	1.96
97.5th Percentile	10.73	8.66	5.37	5.04	4.83	4.69
Increase in P025-P975 range	3.73	2.80	1.29	1.15	1.07	1.00
Rajan-Zingales (1998)						
(Table 4, Column 6)						
T=1067						
IV Slope = 0.31						
IV Standard Error = 0.08						
95% C.I. = (0.16, 0.46)						
<i>Posterior Distribution for Slope</i>						
2.5th Percentile	-1.27	-0.82	-0.06	0.02	0.10	0.16
Mode	0.30	0.31	0.31	0.31	0.31	0.31
97.5th Percentile	1.85	1.41	0.70	0.60	0.53	0.47
Increase in P025-P975 range	10.14	7.26	2.45	1.90	1.40	1.00

Figure 1: The Prior Distribution for $\phi = \text{CORR}(z,u)$

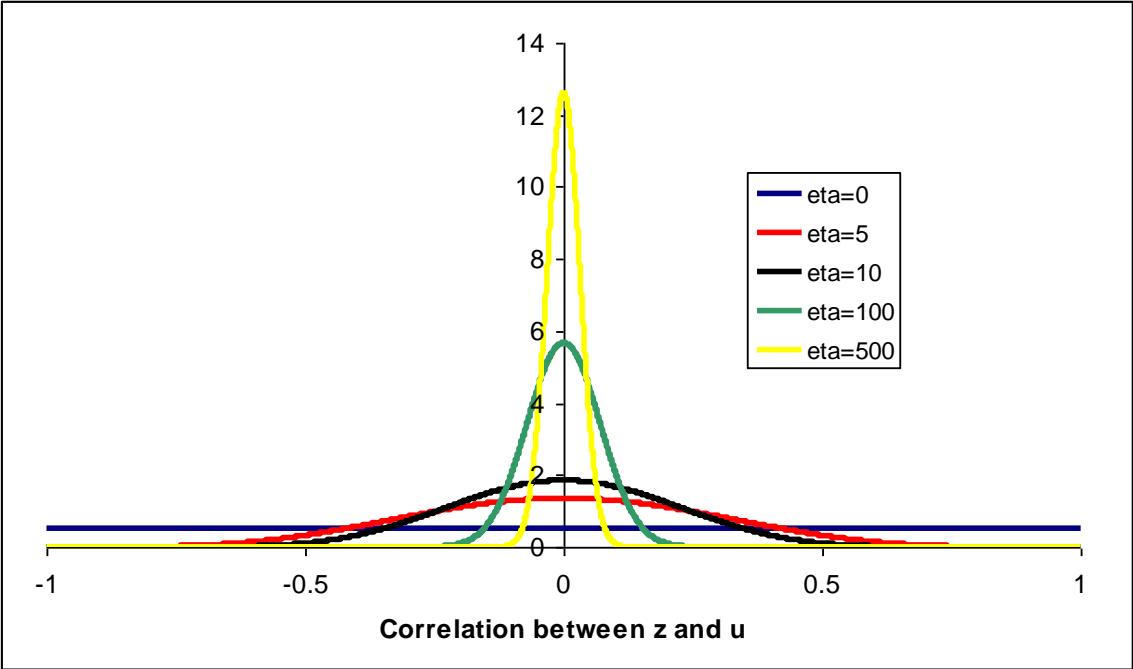


Figure 2: Posterior Distribution for Structural Slopes

