Tractable Bayesian inference for an unidentified simple linear regression model

Robert Calvert Jump, March 18, 2024

University of Greenwich: r.g.calvertjump@greenwich.ac.uk.

Abstract: In this paper, I propose a tractable approach to Bayesian inference in a simple linear regression model for which the standard exogeneity assumption does not hold. By specifying a beta prior for the squared correlation between an error term and regressor, I demonstrate that the implied prior for a bias parameter is t-distributed. If the posterior distribution for the identified regression coefficient is normal, this implies that the posterior distribution for the unidentified treatment effect is the convolution of a normal distribution and a t-distribution. This result is closely related to the literatures on unidentified regression models, imperfect instrumental variables, and sensitivity analysis.

1 Introduction

Consider the simple linear regression model,

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

in which the regressand Y, the regressor X, and the error term U are mean-zero random variables.

As is well known, this model is identified if $\mathsf{E}[U|X] = 0$, which is a sufficient condition for $\mathsf{E}[U] = 0$ and $\mathsf{Cov}[X, U] = 0$ (Wooldridge, 2010, chapter 4). Identification, in this context, means that each unique value of

β implies a unique likelihood function. More precisely, there are no pairs β_1 and β_2 that generate likelihood functions L such that $\mathsf{L}(\beta_1) = \mathsf{L}(\beta_2)$. Identified models, in other words, do not suffer from observational equivalence (Wechsler et al., 2013).

If $\mathsf{E}[U|X] \neq 0$, however, a sample of observations on Y and X will not permit simultaneous estimation of β and $\mathsf{E}[U|X]$. In this case, the linear regression model is unidentified. As discussed below, this is because there exist pairs $(\beta_1, \mathsf{E}[U|X]_1)$ and $(\beta_2, \mathsf{E}[U|X]_2)$ that generate likelihood functions L such that $\mathsf{L}(\beta_1, \mathsf{E}[U|X]_1) = \mathsf{L}(\beta_2, \mathsf{E}[U|X]_2)$.

Unidentified models are considered in various approaches to statistics. The literature on sensitivity analysis, for example, extends back to Cornfield et al. (1959). Sensitivity analyses that are particularly relevant to the linear regression model in (1) include Rosenbaum & Rubin (1983), Frank (2000), Imbens (2003), Hosman et al. (2010), Ashley & Parmeter (2015), Kiviet (2016), Oster (2019), and Cinelli & Hazlett (2020).

Sensitivity analyses solve the inferential problem in (1) by assessing how estimates of β are affected by deterministic changes in functions of $\mathsf{E}[U|X]$. An alternative approach is the Bayesian analysis of unidentified models, in which the posterior distribution of β is conditional on a prior for some function of $\mathsf{E}[U|X]$. This is a generalisation of the standard Bayesian approach, in which the (implicit) prior for $\mathsf{E}[U|X]$ is degenerate at $\mathsf{E}[U|X] = 0$.

This type of Bayesian inference has been widely studied in medical statistics and econometrics. Greenland (2005) is a key reference in the former literature, as is the more recent Greenland (2021). A book-length treatment of Bayesian inference for unidentified models is provided in Gustafson (2015), and Burstyn et al. (2019) provide a rare (for this literature) treatment of Bayesian inference in unidentified linear regression models.

The related literature in econometrics dates back to Leamer (1974) and Kadane (1974), with Florens et al. (1985, 1990) providing Bayesian discussions of unidentified models, and Poirier (1998) providing a particularly clear discussion with useful examples. Kraay (2012), Gustafson (2015), Chan & Tobias (2015) and DiTraglia & Garcia-Jimeno (2021) provide Bayesian treatments of imperfect instrumental variable models, which also give rise to identification problems.

There is, therefore, a substantial literature exploring the Bayesian approach to inference in unidentified models. This literature, however, tends to rely on numerical results, and often focuses on special cases (e.g., Burstyn et al., 2019). This is quite natural in applied contexts, but can make it difficult for those unfamiliar with the literature (particularly students) to understand the estimation problem. The major exception to this rule is Leamer (1974), who (in an under-cited paper) proposes a joint-normal prior for the model in (1) to arrive at a tractable posterior for β.

In this paper, I build on the approach of Leamer (1974) to provide a tractable approach to Bayesian inference in an unidentified simple linear regression model. Specifically, I show that a beta prior for the squared correlation between the error term and regressor in (1) implies a Student's t prior for a bias parameter recently derived in Cinelli & Hazlett (2020). A special case implies a uniform prior for the correlation between the error term and regressor, which allows one to appeal to the principle of insufficient reason in the absence of useful prior information. If the posterior distribution for the identified regression coefficient is normal, then the posterior distribution for the unidentified β is the convolution of a normal distribution and a t-distribution.

Alternatively, the presence of useful prior information can motivate tighter priors for the correlation between the error term and regressor, leading to an inferential problem similar to Leamer (1974). Whether or not this is possible will depend on the relationship being modelled, and will rely on relevant subject knowledge. I provide a simple example, using social science data from the recent extreme bounds analysis in Frank & Martínez i Coma (2023), to illustrate how the approach might be used in practice.

Throughout the paper, upper case Roman letters refer to observable and unobservable variables when considered as random variables, and lower case Roman letters refer to realisations. Following the notational convention in Gustafson (2015), lower case Greek letters with star superscripts refer to parameters when considered as random variables, and lower case Greek letters without superscripts refer to realisations. I refer to the target of inference as a treatment effect, in keeping with the potential outcomes framework, but the approach is valid whenever a linear regression coefficient is consistent for some parameter of interest when regressors are uncorrelated with an error term, and inconsistent otherwise.

2   The model

Instead of relying on the standard exogeneity assumption to justify an estimator for β in (1), assume instead the more general restriction,

$$\mathsf{E}[U|X] = \gamma X. \tag{2}$$

This implies that,

$$Y = \delta X + V, \tag{3}$$

in which $\delta = \beta + \gamma$, $\mathsf{E}[V|X] = 0$, and $\mathsf{E}[Y|X] = \delta X$. One can see that $\delta = \beta + \gamma$ is identified, as distinct values of δ give rise to distinct conditional expectation functions. But the treatment effect β and the bias parameter γ are unidentified, because infinitely many pairs of β and γ give rise to each distinct value of δ. The model described by (1) and (2) is, therefore, unidentified by construction. However, equation (8) in

Cinelli & Hazlett (2020) implies that,

$$|\gamma| = \frac{\sigma_V}{\sigma_X}\sqrt{\frac{\psi}{1-\psi}}, \tag{4}$$

in which $V = Y - \delta X$ is the population analogue of the ordinary least squares residual from (3), σ_V is the standard deviation of V, σ_X is the standard deviation of X, and ψ is the squared correlation between X and U. The result in (4) is derived in online appendix A. It is useful because it isolates ψ, which is the unidentified part of the model.

Motivating a prior for δ is a thoroughly studied problem in Bayesian statistics. But how should one motivate a prior for γ? With limited background knowledge, one might appeal to the principle of insufficient reason, sometimes known as the principle of indifference. This principle states that, if one has no more reason to believe A than B, then one ought not to believe A more than B (Novack, 2010). In other words, where one does not have sufficient reason to regard one possible case as more probable than another, one ought to treat each case as equally probable (Dubs, 1942). For example, in their discussion of unidentified models, Moon & Schorfheide (2012) suggest that, "a prior that is approximately uniform on the identified set for the parameter of interest might serve as a useful benchmark."

The appeal to the principle of insufficient reason in Moon & Schorfheide (2012) removes the need to choose arbitrary parameters for prior distributions on unidentified parameters. However, as ψ can take any value between 0 and 1, the bias in (4) can take any value on the real number line. One cannot, therefore, appeal to the principle of insufficient reason to motivate a (proper) uniform prior for the bias parameter itself.

One can, however, appeal to the principle of insufficient reason to motivate a uniform prior distribution for the correlation between X and U, denoted by ρ. Although this is not the parameter of interest, it is directly related to the parameter of interest and is the underlying source of bias. Moreover, uniform priors have been recommended for correlation coefficients in the identified case (e.g., in Jeffreys, 1961), and there is no obvious reason to suppose that identifiability should influence our choice of prior. Assuming, therefore, that the researcher can motivate the uniform prior,

$$\rho^\star \sim \mathsf{Unif}[-1, 1], \tag{5}$$

we have that $\psi^\star \sim p(\psi) = 0.5\psi^{-0.5}$, or,

$$\psi^\star \sim \mathsf{Beta}(0.5, 1). \tag{6}$$

The prior in (6) suggests the generalisation,

$$\psi^\star \sim \mathsf{Beta}(0.5, 0.5\nu), \tag{7}$$

which leads to a flexible class of prior distributions for γ. To see this, some results for the generalised beta, generalised beta prime, and half-t distributions are briefly recapped.

3 The generalised beta, beta prime, and half-t distributions

Consider a beta-distributed variable Q with parameters κ and λ; $Q \sim \mathsf{Beta}(\kappa, \lambda)$. According to Crooks (2019), it is the case that,

$$T = Q^{1/\omega} \sim \mathsf{GenBeta}(\kappa, \lambda, \omega),$$

in which the generalised beta distribution, $\mathsf{GenBeta}(\kappa, \lambda, \omega)$, has the probability density function,

$$p(t) = \frac{|\omega|}{\mathsf{B}(\kappa, \lambda)} t^{\kappa\omega - 1} (1 - t^\omega)^{\lambda - 1}.$$

If $\kappa = 0.5$, $\lambda = 0.5\nu$, and $\omega = 2$, this expression simplifies to,

$$p(t) \propto (1 - t^2)^{0.5\nu - 1},$$

which is related to the distributions discussed in Berger & Sun (2008) and Ly et al. (2018). Using the same beta-distributed Q, it is also the case that,

$$Z = \alpha + \eta \left(\frac{Q}{1 - Q} \right)^{1/\mu} \sim \mathsf{GenBetaPrime}(\alpha, \eta, \kappa, \lambda, \mu),$$

in which the generalised beta prime distribution, $\mathsf{GenBetaPrime}(\alpha, \eta, \kappa, \lambda, \mu)$, has the probability density function,

$$p(z) = \frac{1}{\mathsf{B}(\kappa, \lambda)} \left| \frac{\mu}{\eta} \right| \left(\frac{z - \alpha}{\eta} \right)^{\kappa\mu - 1} \left(1 + \left(\frac{z - \alpha}{\eta} \right)^\mu \right)^{-\kappa - \lambda}.$$

If $\alpha = 0$, $\kappa = 0.5$, $\lambda = 0.5\nu$, $\mu = 2$, and $\eta = \theta\sqrt{\nu}$, this expression simplifies to,

$$p(z) \propto \left(1 + \frac{1}{\nu} \left(\frac{z}{\theta} \right)^2 \right)^{-(\nu + 1)/2},$$

which is a half-t distribution with ν degrees of freedom (Gelman, 2006; Crooks, 2019).

4 Prior distributions for γ

The results in sections 2 and 3 imply that, if a researcher can motivate a beta prior for ψ, then their prior for $|\rho|$ is a generalised beta distribution and their prior for γ is the non-standardised t-distribution,

$$\gamma^\star | \sigma_V, \sigma_X \sim \mathsf{St}\left(0, \frac{\sigma_V}{\sigma_X \sqrt{\nu}}, \nu \right). \tag{8}$$
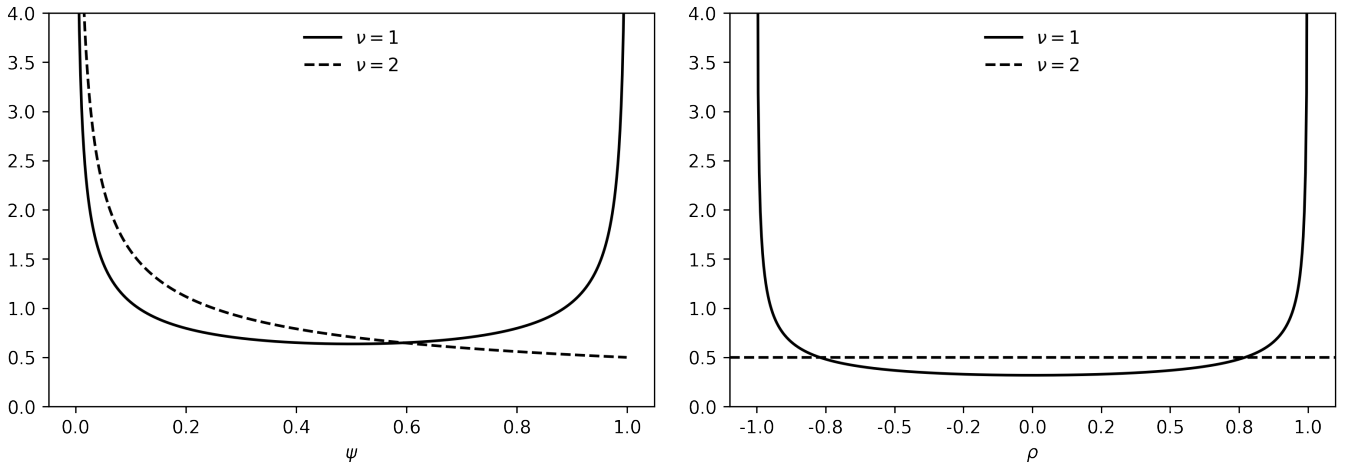
Figure 1: Example priors for ψ (left panel) and ρ (right panel) for two choices of ν in (7).

Figure 1 plots examples of priors for ψ and ρ for two obvious choices of ν. For $\nu = 1$, the prior for ρ resembles (but is not identical to) the prior suggested in Lindley (1965), and the implied prior for γ is a Cauchy distribution (Crooks, 2019). For $\nu = 2$, the prior for ρ is uniform, as in (5). As ν becomes larger, the prior for ρ becomes narrower and the prior for γ becomes approximately normal.

Of course, it would be reasonable to object to privileging the correlation coefficient in this derivation of a prior distribution for γ. Instead, one could place a prior directly on γ, as in Leamer (1974). In fact, the approach outlined above was initially motivated as a way of justifying Leamer's choice. In addition, it is worth noting that the generalised beta prior for ρ has the same form as many of the reference priors in Berger & Sun (2008), with the correspondence depending on the choice of ν.

5 Posterior distributions for β

The model in section 2 is only specified at the level of conditional expectation functions. To provide a more concrete guide to its applicability, consider the following heirarchical representation of a model that can be estimated in practice:

$$
p(Y|\delta, \sigma_V^2, x) = \mathsf{N}\left(\delta x, \sigma_V^2\right),
$$
$$
\delta = \beta + \gamma; \quad p(\gamma^\star|\sigma_V^2, x) = \mathsf{St}\left(0, \frac{\sigma_V}{\hat{\sigma}_X \sqrt{\nu}}, \nu\right),
$$
$$
p(\delta^\star, \sigma_V^{2\star}|x) \propto 1/\sigma_V^2. \tag{9}
$$

The data are now assumed to be normally distributed, and both the data and priors are conditional on the sample values of the regressor. The prior distribution for the estimated regression parameters is uniform on $(\delta, \log\sigma_V)$, i.e., uninformative (Gelman et al., 1995, chapter 14).

Given the foregoing, the conditional posterior distribution for δ is normal,

$$\delta^\star | \sigma_V^2, y, x \sim \mathsf{N}\left(\hat{\delta}, \frac{\sigma_V^2}{s_{xx}}\right), \tag{10}$$

in which $\hat{\delta}$ is the OLS estimator and s_{xx} is the sample sum of squares of the regressor, while the posterior for σ_V^2 is a scaled inverse-χ^2 distribution,

$$\sigma_V^{2\star} | y, x \sim \mathsf{Inv\text{-}}\chi^2\left(n - 1, \hat{\sigma}_V^2\right), \tag{11}$$

in which $\hat{\sigma}_V^2$ is the sample standard deviation of the OLS residuals.

As $\delta = \beta + \gamma$, the conditional posterior distribution for β is a convolution of the normal conditional posterior for δ and the Student's t prior for γ,

$$\begin{aligned} \beta^\star | \sigma_V, y, x &= \delta^\star | \sigma_V, y, x \\ &- \gamma^\star | \sigma_V, x. \end{aligned} \tag{12}$$

As described in Pogány & Nadarajah (2013), the probability density function of this convolution can be expressed as,

$$p(\beta) \propto \exp\left(-\frac{\left(\beta - \hat{\delta}\right)^2}{2\frac{\sigma_V^2}{s_{xx}}}\right) \times g(\beta),$$

in which $g(\beta)$ involves an integral that is explained in online appendix B, and is straightforward to evaluate numerically.

To compute the marginal posterior distribution for β, σ_V^2 needs to be integrated out of the conditional posteriors for δ and β in (12), using (11). This is straightforward to achieve using simulations, although certain choices of ν in the prior for γ might imply the need for a large number of runs to accurately estimate the tails in $\gamma^\star | x$.

On the other hand, if both ν and the sample size are large enough, the marginal posteriors for both δ and γ will be approximately normal, and so too will the marginal posterior for β. These issues are considered further in the next section.

6   Practical considerations

The model outlined in section 5 is simplified to permit tractability, in order to aid pedagogy. However, if one can motivate a reasonable value for ν, then the model can be estimated. Moreover, the prior for (δ, σ_V) could be changed without major complication (e.g., by using a conjugate prior).

How, then, should a researcher motivate a choice for ν? This will depend on the relationship being modelled and the presence of expert subject knowledge. But progress can be made by examining existing extreme bounds analyses in the relevant subject, to give some idea about the correlation structure of commonly used covariates (Leamer, 1983).

Consider, for example, the problem of identifying the determinants of voter turnout in elections. This is of obvious policy importance, as the overall health of a democracy is closely related to the engagement of its citizens in the electoral process. Despite this, there is very little consensus in the political science literature on these determinants, mainly due to a lack of convincing identification strategies. One obvious possibility is the cross-country relationship between voter turnout and income inequality,

$$Y = \beta X + U, \tag{13}$$

in which Y denotes the number of voters in an election as a percentage of the voting population, and X denotes the Gini coefficient of income inequality. Clearly, $\mathsf{E}[U|X] = \gamma X$ in the general case.

This problem motivates the recent extreme bounds analysis in Frank & Martínez i Coma (2023). They estimate repeated linear regression models of the form described in (13), augmented with,

$$U = \zeta C + E, \tag{14}$$

in which C is a vector of different combinations of over 60 observable covariates, and E is an error term.

Using their dataset, the left panel of figure 2 plots the empirical distributions of the squared correlation ψ between X and different choices of U, calculated from approximately 37,000 regressions (further details of the computations are provided in online appendix C). This histogram is overlaid with the beta prior in (7), in which $\nu = 134$.

In this example, the empirical distribution of ψ is relatively tight. Moreover, the residual degrees of freedom in the simple linear regression of Y on X is 397, so one can treat the marginal posterior distributions of both δ and β as approximately normal. The posterior mean and variance for δ, using the model in section 5, are -1.08 and 0.46, while the posterior for β is approximated by,

$$\beta^\star | y, x \sim \mathsf{N}\left(\hat{\delta}, \frac{\hat{\sigma}_V^2}{s_{xx}} + \frac{\hat{\sigma}_V^2}{\hat{\sigma}_X^2 \nu}\right), \tag{15}$$

which has a mean and variance of -1.08 and 1.83. Approximately 79% of this posterior mass is for negative values of β, with the remaining 21% positive.
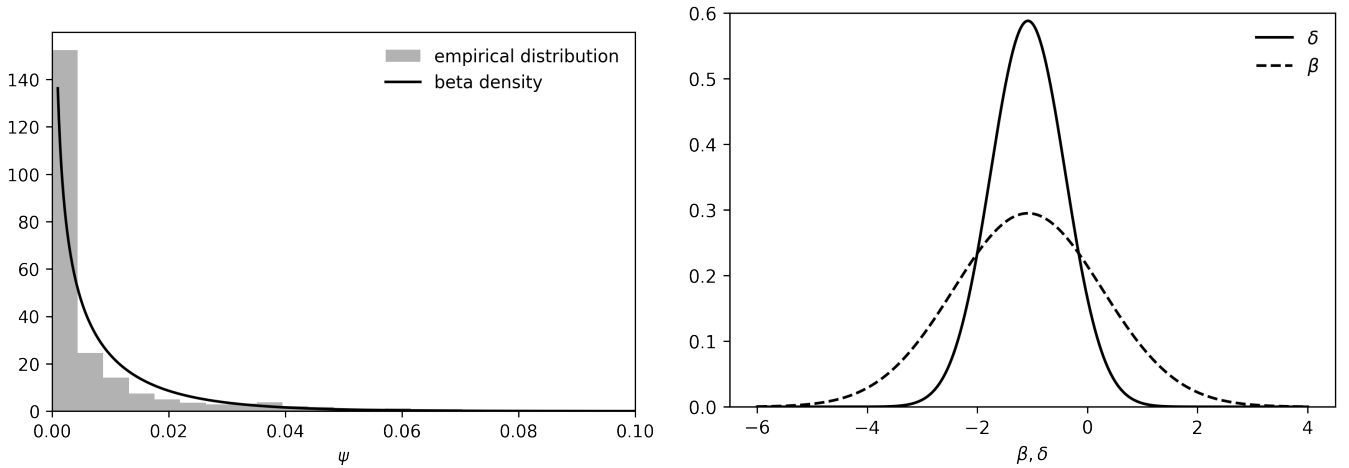
Figure 2: Empirical distribution of ψ from Frank & Martínez i Coma (2023) (left panel) and normal approximations to the posterior distributions for δ and β (right panel) in (13).

These normal approximations to the posterior distributions of δ and β are displayed in the right panel of figure 2. The increased uncertainty surrounding the treatment effect β compared to the reduced form δ is readily apparent, which is qualitatively similar to the results in Frank & Martínez i Coma (2023), in which the extreme bounds distribution of the coefficient on income inequality varies widely across model specifications.

The normal approximation in (15) illustrates in a particularly clear manner that the posterior standard deviation of β shrinks to a positive constant as the sample size increases, rather than zero. In general, as discussed Gustafson (2005), Moon & Schorfheide (2012), Gustafson (2015), and others, the posterior of the identified δ tends to the true value of δ as the sample size increases, while the posterior of the unidentified β tends to its conditional prior evaluated at the true value of δ.

Of course, in other cases a choice of large ν might be inappropriate. Moreover, one might be concerned that, even though scores of observable confounders have been used to estimate the distribution of ψ in this example, the remaining unobservable confounders might follow a different distribution. In these cases researchers could utilise a uniform prior for ρ, appealing to the principle of insufficient reason, or even the extreme prior obtained by setting $\nu = 1$, which implies that the posterior for β is the convolution of a normal distribution and a Cauchy distribution, which is a Voigt distribution (Crooks, 2019).

7   Discussion

According to the data visualisation pioneer Edward Tufte, one of the "shortest true statements" that can be made about statistics is that, "correlation is not causation but it sure is a hint" (Tufte, 2006). This outlook informs various approaches to statistics that attempt to deal with imperfect identification, including the literatures on unidentified regression models, imperfect instrumental variables, and sensitivity analysis.

In this paper, I hope to have contributed to these literatures by providing a tractable approach to Bayesian inference for unidentified linear regression models. Although the estimation problem discussed in this paper does not disappear as the sample size increases, it does disappear as the correlation between the regressor and regressand becomes stronger, and σ_V approaches zero in (4). Simply put, the stronger the correlation between the regressor and regressand, the greater the likelihood of a positive treatment effect.

The results in this paper are most closely related to those in Leamer (1974) and Cinelli & Hazlett (2020). In fact, thinking about how one might appeal to the principle of insufficient reason to motivate a choice of variance for the bias prior in Leamer (1974) was the original inspiration behind this paper, and was made possible by the derivations in Cinelli & Hazlett (2020). Various contributions in the medical statistics literature, including Gustafson (2005, 2009, 2010, 2012, 2014), McCandless et al. (2007, 2008), Gustafson et al. (2010), Xia & Gustafson (2016), and Gustafson & McCandless (2018), are also closely related.

Importantly, the Bayesian approach should not be seen as a competitor to sensitivity analyses, but rather a complement. In fact, one could conduct a sensitivity analysis on the Bayesian approach outlined in this paper, by observing how the posterior for β changes when ν is varied deterministically. At the same time, using prior distributions can simply the inferential procedure somewhat, by reducing the amount of parameters that have to be varied deterministically.[1]

Finally, there are many ways that the approach outlined in this paper could be extended to make it less focused on simplicity for pedagogical reasons, and more focused on practical application. One obvious avenue is an extension to multiple linear regressions with observable controls, perhaps borrowing from the frequentist approach in Oster (2019). Another would be an exploration of the correlation structure of observable covariates using other extreme bounds analyses, or novel databases.

In any case, I hope that the tractability of the results presented here will help those unfamiliar with the literature to understand the peculiar problems of estimation and inference in unidentified models, and apply simple solutions in practice.

Conflict of interest statement

---

[1]The argument that sensitivity analysis should be agnostic to inferential procedure was originally put to me by Carlos Cinelli, in private correspondence.

References

Ashley, R. A., & Parmeter, C. F. (2015). When is it justifiable to ignore explanatory variable endogeneity in a regression model? *Economics Letters*, *137*, 70–74.

Berger, J. O., & Sun, D. (2008). Objective priors for the bivariate normal model. *The Annals of Statistics*, *36*, 963–982.

Burstyn, I., Barone-Adesi, F., De Vocht, F., & Gustafson, P. (2019). What to do when accumulated exposure affects health but only its duration was measured? A case of linear regression. *International Journal of Environmental Research and Public Health*, *16*(11), 1896.

Chan, J. C., & Tobias, J. L. (2015). Priors and posterior computation in linear endogenous variable models with imperfect instruments. *Journal of Applied Econometrics*, *30*(4), 650–674.

Cinelli, C., & Hazlett, C. (2020). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *82*(1), 39–67.

Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., & Wynder, E. L. (1959). Smoking and Lung Cancer: Recent Evidence and a Fiscussion of Some Wuestions. *Journal of the National Cancer institute*, *22*(1), 173–203.

Crooks, G. E. (2019). Field Guide to Continuous Probability Distributions. Tech. rep., Berkeley Institute for Theoretical Sciences.
URL http://threeplusone.com/fieldguide

DiTraglia, F. J., & Garcia-Jimeno, C. (2021). A framework for eliciting, incorporating, and disciplining identification beliefs in linear models. *Journal of Business & Economic Statistics*, *39*(4), 1038–1053.

Dubs, H. H. (1942). The principle of insufficient reason. *Philosophy of Science*, *9*(2), 123–131.

Florens, J., Mouchart, M., & Rolin, J. (1985). On two definitions of identification. *Statistics: A Journal of Theoretical and Applied Statistics*, *16*(2), 213–218.

Florens, J., Mouchart, M., & Rolin, J. (1990). Elements of Bayesian statistics. *Monographs and Textbooks in Pure and Applied Mathematics*.

Frank, K. A. (2000). Impact of a confounding variable on a regression coefficient. *Sociological Methods & Research*, *29*(2), 147–194.

Frank, R. W., & Martínez i Coma, F. (2023). Correlates of voter turnout. *Political behavior*, *45*(2), 607–633.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian Analysis*, *1*(3), 515–534.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian Data Analysis*. Chapman and Hall/CRC.
URL http://www.stat.columbia.edu/~gelman/book/

Greenland, S. (2005). Multiple-bias modelling for analysis of observational data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *168*(2), 267–306.

Greenland, S. (2021). Invited commentary: dealing with the inevitable deficiencies of bias analysis - and all analyses. *American Journal of Epidemiology*, *190*(8), 1617–1621.

Gustafson, P. (2005). On model expansion, model contraction, identifiability and prior information: two illustrative scenarios involving mismeasured variables. *Statistical Science*, *20*(2), 111–140.

Gustafson, P. (2009). What are the limits of posterior distributions arising from nonidentified models, and why should we care? *Journal of the American Statistical Association*, *104*(488), 1682–1695.

Gustafson, P. (2010). Bayesian inference for partially identified models. *The International Journal of Biostatistics*, *6*(2).

Gustafson, P. (2012). On the behaviour of Bayesian credible intervals in partially identified models. *Electronic Journal of Statistics*, *6*, 2107–2124.

Gustafson, P. (2014). Bayesian inference in partially identified models: Is the shape of the posterior distribution useful? *Electronic Journal of Statistics*, *8*(1), 476–496.

Gustafson, P. (2015). *Bayesian inference for partially identified models: Exploring the limits of limited data*, vol. 140. CRC Press.

Gustafson, P., McCandless, L., Levy, A., & Richardson, S. (2010). Simplified Bayesian Sensitivity analysis for mismeasured and unobserved confounders. *Biometrics*, *66*(4), 1129–1137.

Gustafson, P., & McCandless, L. C. (2018). When is a sensitivity parameter exactly that? *Statistical Science*, *33*(1), 86–95.

Hosman, C. A., Hansen, B. B., Holland, P. W., et al. (2010). The sensitivity of linear regression coefficients' confidence limits to the omission of a confounder. *The Annals of Applied Statistics*, *4*(2), 849–870.

Imbens, G. W. (2003). Sensitivity to exogeneity assumptions in program evaluation. *American Economic Review*, *93*(2), 126–132.

Jeffreys, H. (1961). *Theory of Probability, 3rd Edition*. The Clarendon Press, Oxford University Press, New York.

Kadane, J. (1974). The role of identification in Bayesian theory. In S. Fienberg, & A. Zellner (Eds.) *Bayesian Econometrics and Statistics*. North Holland.

Kiviet, J. F. (2016). When is it really justifiable to ignore explanatory variable endogeneity in a regression model? *Economics Letters*, *145*, 192–195.

Kraay, A. (2012). Instrumental variables regressions with uncertain exclusion restrictions: a Bayesian approach. *Journal of Applied Econometrics*, *27*(1), 108–128.

Leamer, E. (1974). False Models and Post-Data Model Construction. *Journal of the American Statistical Association*, *69*(345), 122–131.

Leamer, E. E. (1983). Let's take the con out of econometrics. *The American Economic Review*, *73*(1), 31–43.

Lindley, D. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint (Part 2: Inference)*. Cambridge University Press.

Ly, A., Marsman, M., & Wagenmakers, E.-J. (2018). Analytic posteriors for Pearson's correlation coefficient. *Statistica Neerlandica*, *72*(1), 4–13.

McCandless, L. C., Gustafson, P., & Levy, A. (2007). Bayesian sensitivity analysis for unmeasured confounding in observational studies. *Statistics in Medicine*, *26*(11), 2331–2347.

McCandless, L. C., Gustafson, P., & Levy, A. (2008). A sensitivity analysis using information about measured confounders yielded improved uncertainty assessments for unmeasured confounding. *Journal of Clinical Epidemiology*, *61*(3), 247–255.

Moon, H. R., & Schorfheide, F. (2012). Bayesian and frequentist inference in partially identified models. *Econometrica*, *80*(2), 755–782.

Novack, G. (2010). A Defense of the Principle of Indifference. *Journal of Philosophical Logic*, *39*, 655–678.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, *37*(2), 187–204.

Pogány, T. K., & Nadarajah, S. (2013). On the convolution of normal and t random variables. *Statistics*, *47*(6), 1363–1369.

Poirier, D. (1998). Revising beliefs in nonidentified models. *Econometric Theory*, *14*(4), 483–509.

Rosenbaum, P. R., & Rubin, D. B. (1983). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society: Series B (Methodological)*, *45*(2), 212–218.

Tufte, E. (2006). *The Cognitive Style of PowerPoint: Pitching Out Corrupts Within*. Graphics Press LLC.

Wechsler, S., Izbicki, R., & Esteves, L. G. (2013). A Bayesian look at nonidentifiability: A simple example. *The American Statistician*, *67*(2), 90–93.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT Press.

Xia, M., & Gustafson, P. (2016). Bayesian regression models adjusting for unidirectional covariate misclassification. *Canadian Journal of Statistics*, *44*(2), 198–218.

Online appendices for 'Tractable Bayesian inference for an unidentified simple linear regression model'

A Deriving the bias parameter

As discussed in section 2 of the main text, the bias parameter in (4) is implied by equation (8) in Cinelli & Hazlett (2020). They consider a data generating process of the form,

$$Y = \beta X + \zeta U + W, \tag{A.1}$$

which nests the model in (1), and an estimated (restricted) model of the form,

$$Y = \delta X + V. \tag{A.2}$$

Starting from the standard omitted variable bias formula, we have,

$$
\begin{aligned}
\mathsf{bias} = \gamma = \delta - \beta &= \frac{\mathsf{Cov}(X,U)}{\mathsf{Var}(X)} \times \zeta, \\
&= \frac{\mathsf{Cov}(X,U)}{\mathsf{Var}(X)} \times \frac{\mathsf{Cov}(Y^{\perp X}, U^{\perp X})}{\mathsf{Var}(U^{\perp X})}, \\
&= \frac{\mathsf{Cor}(X,U)\mathsf{Sd}(U)}{\mathsf{Sd}(X)} \times \frac{\mathsf{Cor}(Y^{\perp X}, U^{\perp X})\mathsf{Sd}(Y^{\perp X})}{\mathsf{Sd}(U^{\perp X})}, \\
&= \frac{\mathsf{Cor}(X,U)\mathsf{Cor}(Y^{\perp X}, U^{\perp X})}{\frac{\mathsf{Sd}(U^{\perp X})}{\mathsf{Sd}(U)}} \times \frac{\mathsf{Sd}(Y^{\perp X})}{\mathsf{Sd}(X)},
\end{aligned}
\tag{A.3}
$$

in which $Y^{\perp X}$ is the variable Y after removing the component linearly explained by X, and similarly for $U^{\perp X}$. Now, recalling that $\frac{\mathsf{Var}(U^{\perp X})}{\mathsf{Var}(U)} = 1 - \mathsf{R}^2_{UX} = 1 - \mathsf{R}^2_{XU}$, in which $\mathsf{R}^2_{XU} = \mathsf{R}^2_{UX}$ is the R^2 from regressing X on U, we have,

$$|\gamma| = \sqrt{\frac{\mathsf{R}^2_{XU}\mathsf{R}^2_{YU|X}}{1 - \mathsf{R}^2_{XU}}} \times \frac{\mathsf{Sd}(Y^{\perp X})}{\mathsf{Sd}(X)}, \tag{A.4}$$

in which $\mathsf{R}^2_{YU|X}$ is the partial R^2 from regressing Y on U after controlling for X. If $\mathsf{Var}(W) = 0$ then the model in (A.1) is equivalent to the model in (1), noting that the value of ζ is irrelevant for the (hypothetical) unobservable U. In which case, $\mathsf{R}^2_{YU|X} = 1$ and thus,

$$|\gamma| = \sqrt{\frac{\mathsf{R}^2_{XU}}{1 - \mathsf{R}^2_{XU}}} \times \frac{\mathsf{Sd}(Y^{\perp X})}{\mathsf{Sd}(X)}, \tag{A.5}$$

which is equivalent to (4) with a change in notation.

B   The convolution of a normal distribution and a t-distribution

Using the notation in Pogány & Nadarajah (2013), consider a normally distributed X and t-distributed Y such that,

$$X \sim p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \tag{B.1}$$

$$Y \sim p(y) = \frac{1}{\sqrt{\nu}\lambda \mathsf{B}(\nu/2, 1/2)} \left[1 + \frac{(y-m)^2}{\nu\lambda^2}\right]^{-(\nu+1)/2}. \tag{B.2}$$

Equations (3) and (4) in Pogány & Nadarajah (2013) state that the probability density function of $S = X+Y$ can be expressed as,

$$p(s) = \frac{1}{\sqrt{2\pi}\sigma \mathsf{B}(\nu/2, 1/2)} \exp\left[-\frac{(s-\mu-m)^2}{2\sigma^2}\right] \mathsf{I}\left(\frac{\lambda^2\nu}{2\sigma^2}, \frac{\lambda\sqrt{\nu}(\mu+m-s)}{\sigma^2}, \frac{\nu+1}{2}\right), \tag{B.3}$$

in which,

$$\mathsf{I}(a,b,c) = \int_{-\infty}^{\infty} \frac{\exp(-au^2 - bu)}{(1+u^2)^c}\,\mathrm{d}u. \tag{B.4}$$

Translating from this notation to the notation of the model outlined in the main text, $s = \beta$, $\mu = \hat{\delta}$, $m = 0$, $\sigma^2 = \sigma_V^2/s_{xx}$, and $\lambda = \sigma_V/(\hat{\sigma}_X\sqrt{\nu})$. This convolution can be computed using the `mpmath` package in Python with the following code snippet:

```
1  # grid for s:
2  size = 100
3  s_space = mp.linspace(-5, 5, size)
4  # create array to hold pdf
5  pdf = mp.zeros(size, 1)
6  # set model parameters:
7  mu_ = . . . . mu = mp.mpf(mu_) . . . etc
8  # composite constants:
9  a = mp.power(lam,2)*v/(2*mp.power(sig,2))
10 c = mp.mpf((v+1)/2)
11 const = 1 / ( mp.sqrt(2*mp.pi)*sig*mp.beta(v/2,1/2) )
12
13 # compute pdf
14 for idx, s in enumerate(s_space):
15
16     # compute composite parameter b
17     b = lam*mp.sqrt(v)*(mu + m - s)
18
19     # compute integral
20     f = lambda u: mp.exp(-a*u**2 - b*u) / mp.power((1 + mp.power(u,2)), c)
21     integral = mp.quad(f, [-mp.inf, mp.inf])
22
23     # compute pdf
24     pdf[idx] = const*mp.exp(-mp.power(s - mu - m, 2)/(2*mp.power(sig,2)))*integral
```

C The empirical example

The empirical example in section 6 in the main text uses the dataset from Frank & Martínez i Coma (2023) downloaded from,

on 12th February 2024. The dependent variable is voter turnout as a percentage of voting age population, and the covariates are the 69 variables listed in table A12 of that paper's appendix. Excluding the Gini coefficient, that leaves 68 potential confounders. The potential variables in C are then made up of the list of all subsets of these 68 potential confounders of size three or smaller, which implies 52,462 separate regressions. After excluding regressions with multicollinearity problems and regressions with highly consequential missing observations, the total number of regressions run is 36,698.

Running these regressions yields 36,698 separate 'draws' of ψ, the squared correlation between X and U, from the empirical distribution of observable confounders. As the distribution of ψ is assumed to be $\mathsf{Beta}(0.5, 0.5\nu)$, this implies $\mathsf{E}[\psi] = 1/(1 + \nu)$, which motivates a simple moment estimator of ν,

$$\hat{\nu} = \frac{1 - \hat{\mathsf{E}}[\psi]}{\hat{\mathsf{E}}[\psi]} = 133.54. \tag{B.5}$$

Of course, as noted in the main text, a choice of ν this large might be inappropriate in other contexts. Moreover, one might be concerned that, even though scores of observable confounders have been used to estimate the distribution of ψ in this example, the remaining unobservable confounders might follow a different distribution. In these cases, as already noted, researchers could appeal to the principle of insufficient reason to motivate a uniform prior for ρ.