



Priors and Posterior Computation in Linear Endogenous Variable Models with Imperfect Instruments

Joshua C.C. Chan
Australian National University

Justin L. Tobias
Purdue University

ANU Working Papers in Economics and Econometrics
580

August 2012

JEL: C11, I10, J11

ISBN: 0 86831 580 X

PRIORS AND POSTERIOR COMPUTATION IN LINEAR ENDOGENOUS VARIABLE MODELS WITH IMPERFECT INSTRUMENTS

JOSHUA C.C. CHAN AND JUSTIN L. TOBIAS

AUSTRALIAN NATIONAL UNIVERSITY AND PURDUE UNIVERSITY

ABSTRACT.

Estimation in models with endogeneity concerns typically begins by searching for instruments. This search is inherently subjective and identification is generally achieved upon imposing the researcher's strong prior belief that such variables have no conditional impacts on the outcome. Results obtained from such analyses are necessarily conditioned upon the untestable opinions of the researcher, and such beliefs may not be widely shared.

In this paper we, like several studies in the recent literature, employ a Bayesian approach to estimation and inference in models with endogeneity concerns by imposing weaker prior assumptions than complete excludability. When allowing for instrument imperfection of this type, the model is only partially identified, and as a consequence, standard estimates obtained from the Gibbs simulations can be unacceptably imprecise. We thus describe a substantially improved "semi-analytic" method for calculating parameter marginal posteriors of interest that only requires use of the well-mixing simulations associated with the *identifiable* model parameters and the form of the conditional prior. Our methods are also applied in an illustrative application involving the impact of Body Mass Index (BMI) on earnings.

Keywords: Bayes, Endogeneity, Identification, MCMC

JEL classification: C11, I10, J11

Date: July 24, 2012. We thank seminar participants at Virginia Tech for helpful comments and suggestions. All errors are, of course, our own. Email: joshua.chan@anu.edu.au and jltobias@purdue.edu.

1. INTRODUCTION

Practitioners seeking to estimate the impact of an endogenous treatment on an outcome or set of outcomes generally follow a similar research strategy. First and most importantly, a set of variables are found that (a) are believed to be conditionally related to the treatment decision yet (b) can plausibly be excluded from the outcome equation. The first of these two conditions is empirically testable, while the second is untestable. The determination of such variables is so critically important to the process that their availability (or non-availability) often defines the research agenda; when plausible instruments present themselves, research ensues and papers are written that exploit their existence, while interesting economic causal effect questions that lack obvious exclusion restrictions often remain unexplored or under-explored. Second, once a set of variables of this type are determined, a variety of suggestive evidence is presented with the intention of bolstering the case for the instrument's validity. Finally, traditional estimation methods such as MLE, IV or 2SLS estimation are applied, producing a (hopefully consistent) estimate of the causal effect of interest.

To the Bayesian, a view that one has valid instrumental variables is akin to employing a particular dogmatic prior in which the researcher imposes that coefficients associated with these variables have zero values in the outcome equation. These priors are controversial in many / most cases, and attempts to "prove" the appropriateness of such priors, though ubiquitous, are perhaps totally unfounded from the subjectivist's perspective. Furthermore, even seemingly valid, highly influential and often used instruments such as quarter of birth in education studies [e.g., Angrist and Krueger (1991)] have recently come under scrutiny as valid exclusion restrictions [Buckles and Hungerman (2010)], potentially suggesting difficulty in ever finding fully convincing instruments with observational data. Thus, while one might not only question the value in any effort that tries to "prove" the correctness of the prior, even if such exercises were deemed useful, it is not likely that they will ever be completely convincing with non-experimental data.

A goal of this paper, following the interesting work of Conley et al (2012), Kraay (2012) and Nevo and Rosen (2010) among others, is to offer an operational alternative to the dogmatic priors that are commonly employed in empirical work of this type. The priors we employ in our paper are somewhat similar in spirit to those employed in Conley et al (2012), and our paper does not claim to make a substantial contribution in terms of prior specification. Our focus instead is to document the potential for serious inaccuracies when standard Gibbs approaches to posterior simulation are conducted in the presence of imperfect excludability, and to introduce a new and

significantly more efficient computational approach for calculating marginal posteriors of interest and for investigating prior sensitivity.

We formally discuss how learning takes place regarding (non-identified) parameters of our partially identified model and use the lessons learned in that exercise to improve upon standard Gibbs approaches to posterior simulation. Specifically, the lack of point identification leads to very poor mixing of the Gibbs simulations, and this problem is exacerbated as one increases prior uncertainty surrounding the validity of the instruments. We thus develop a “semi-analytic” scheme for the calculation of parameter marginal posterior distributions and their features. This technique only uses the form of the conditional prior and the well-mixing *identified* posterior simulations from the standard Gibbs algorithm to perform the requisite calculations. Given the Rao-Blackwell spirit of the methods we develop, we can even improve upon *iid* sampling from the posterior.

We illustrate our methods using data from a recent study by Kline and Tobias (2008). These authors investigate the causal impact of BMI (Body Mass Index) on earnings, and use parental BMI as exclusion restrictions. We obtain results under the assumption that parental BMI is excludable, and also under the assumption that these variables are not excludable. Our results suggest that, when departing from the dogmatic prior of Kline and Tobias and those commonly used in this literature, we are still able to obtain informative posterior results. Furthermore, we find evidence that the BMI log-wage relationship is negative under a variety of different weak prior beliefs.

The outline of this paper is as follows. The following section presents a stylized model, describes our priors and briefly discusses our posterior simulator under those priors. Section 3 formally describes how learning takes place regarding the structural parameters of the model, illustrating this issue specifically for the causal effect parameter. Section 4 provides an illustrative example investigating the impact of Body Mass Index (BMI) on earnings and compares our methods for posterior calculation with results obtained using traditional methods. The paper concludes with a summary in section 5.

2. THE MODEL AND PRIOR

Our primary model is a two-equation triangular system, with y denoting a continuous outcome variable and s denoting a single continuous (and potentially endogenous) explanatory variable.

The basic specification we consider has the form:¹

$$(1) \quad y_i = \alpha_0 + z_i\alpha_1 + \mathbf{x}_i\boldsymbol{\alpha}_2 + s_i\gamma + u_i,$$

$$(2) \quad s_i = \beta_0 + z_i\beta_1 + \mathbf{x}_i\boldsymbol{\beta}_2 + v_i,$$

where

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} \Big|_{\mathbf{x}_i, \mathbf{z}_i} \stackrel{iid}{\sim} N \left[\mathbf{0}, \begin{pmatrix} \sigma_u^2 & \rho_{uv}\sigma_u\sigma_v \\ \rho_{uv}\sigma_u\sigma_v & \sigma_v^2 \end{pmatrix} \right].$$

The distinction between \mathbf{x} and z in (1) and (2) is that \mathbf{x} represents covariates that are not believed to be potential exclusion restrictions and thus are introduced to affect both the endogenous variable and the conditional outcome. The variable z , like \mathbf{x} , is also included in both equations although there is a reasonably strong prior belief that it could be dropped from (1). Instead of imposing *a priori* that $\alpha_1 = 0$, however, we permit the possibility that there may be some “direct effect” of z on y above and beyond its influence through s . Given our understanding of the problem under study, and its potential excludability from (1) derived from that knowledge, we will differentiate z from \mathbf{x} in the priors that we employ over their respective coefficients.

The triangular system in (1) and (2) is not fully identifiable even under Gaussianity: the model contains ten (sets of) parameters — $\alpha_0, \alpha_1, \boldsymbol{\alpha}_2, \gamma, \beta_0, \beta_1, \boldsymbol{\beta}_2, \rho_{uv}, \sigma_u^2$ and σ_v^2 — but only nine moment conditions that can be exploited for their estimation.² To see this, employ a marginal-conditional decomposition and factor the likelihood $p(y, s|\cdot)$ into $p(y|s, \cdot)$ and $p(s|\cdot)$. Then note that the expectation and variance of s (marginal of y yet conditioned on the covariates and model parameters) are, respectively

$$\mathbb{E}(s|\cdot) = \beta_0 + \beta_1 z + \mathbf{x}\boldsymbol{\beta}_2, \quad \text{Var}(s|\cdot) = \sigma_v^2,$$

which gives four (sets of) moment conditions. In addition, the conditional expectation $\mathbb{E}(y|s)$ and conditional variance $\text{Var}(y|s)$ provide another five:

$$\begin{aligned} \mathbb{E}(y|s, \cdot) &= \left(\alpha_0 - \frac{\rho_{uv}\sigma_u}{\sigma_v}\beta_0 \right) + \left(\alpha_1 - \frac{\rho_{uv}\sigma_u}{\sigma_v}\beta_1 \right) z + \mathbf{x} \left(\boldsymbol{\alpha}_2 - \frac{\rho_{uv}\sigma_u}{\sigma_v}\boldsymbol{\beta}_2 \right) + \left(\gamma + \frac{\rho_{uv}\sigma_u}{\sigma_v} \right) s \\ \text{Var}(y|s, \cdot) &= (1 - \rho_{uv}^2)\sigma_u^2. \end{aligned}$$

¹We fix ideas on the continuous-treatment, continuous-outcome case throughout this paper.

²Since the normal distribution is characterized by its first two moments, given the nine moment conditions we can reconstruct the likelihood. In other words, all the information in the likelihood is summarized by the nine moment conditions.

Hence, the identified parameters are $\boldsymbol{\psi} = (a_0, a_1, \mathbf{a}_2, \beta_0, \beta_1, \boldsymbol{\beta}_2, \sigma_v^2, b, c^2)'$ where

$$a_0 \equiv \alpha_0 - \frac{\rho_{uv}\sigma_u}{\sigma_v}\beta_0, \quad a_1 \equiv \alpha_1 - \frac{\rho_{uv}\sigma_u}{\sigma_v}\beta_1, \quad \mathbf{a}_2 \equiv \boldsymbol{\alpha}_2 - \frac{\rho_{uv}\sigma_u}{\sigma_v}\boldsymbol{\beta}_2, \quad b \equiv \gamma + \frac{\rho_{uv}\sigma_u}{\sigma_v}, \quad c^2 \equiv (1 - \rho_{uv}^2)\sigma_u^2,$$

whereas the structural parameters are $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \boldsymbol{\alpha}_2, \beta_0, \beta_1, \boldsymbol{\beta}_2, \gamma, \sigma_u^2, \sigma_v^2, \rho_{uv})$. In other words, the data are only informative regarding some structural parameters of interest, and certain functions of their values.

2.1. Priors (Implicitly) Employed in the Literature. The usual exclusion restriction path to model identification proceeds by thinking of a z for which the prior $p(\alpha_1) = \mathbf{1}(\alpha_1 = 0)$ can be credibly maintained, yet $\beta_1 \neq 0$. The priors we introduce in this paper relax this dogmatic stance, and are more easily defended in practice, although their use leads us to work with a partially identified model.

One possible restriction one could employ — instead of imposing that $\alpha_1 = 0$ — would instead order the sizes of z 's possible relationships with y : α_1 , which we term the “direct” effect of z on y , (which is commonly imposed to equal zero), should be smaller than the instrument's primary channel of influence, which is through its relationship with the endogenous variable s .

There are a number of ways to impose this type of belief. Such an ordering could be imposed by specifying, for example, a conditional prior of the form:³

$$(3) \quad \alpha_1 | \beta_1, \gamma \sim \mathcal{TN}_{(-|\beta_1\gamma|, |\beta_1\gamma|)}(0, \underline{V}_{\alpha_1}),$$

with $\mathcal{TN}_{(a,b)}(\mu, \sigma^2)$ denoting a normal distribution with mean μ and variance σ^2 that is truncated to the interval (a, b) . Thus, we choose to center α_1 over zero *a priori* (since our knowledge of the problem and identification of z as a candidate instrument suggests that it may be excludable), and also impose that it must be smaller in absolute magnitude than $|\beta_1\gamma|$ — which we term the “indirect effect” — the absolute marginal effect of z on y through its relationship with s .⁴

While this prior seems sensible and asserts far less than the dogmatic $\alpha_1 = 0$, it is important to note that adopting this prior precludes the possibility of finding no causal effect — that is,

³Note that priors like this one, and the others we consider, yield posteriors for which direct analytic solutions are not available. Thus, in general, one will need to use simulation methods to calculate posterior statistics of interest.

⁴In the context of the prominent example mentioned in the introduction, our prior would allow quarter of birth to have some conditional relationship with earnings, but impose that the size of this direct effect must be smaller than indirect effect that arises through the relationships between quarter of birth and educational attainment, and educational attainment and earnings.

0 will not be contained in the support of the marginal posterior of γ .⁵ Given this consequence, we believe that such a prior might be credible to employ in some situations, such as determining the effect of education on earnings (which is widely believed to be non-zero), yet may be rather questionable in others, such as investigating the impact of body mass on earnings (for which existing results are decidedly mixed).

One can sidestep this issue by eliminating γ from the conditioning in (3). To obtain a prior similar to that in (3) but free of γ , one could instead elicit something like a maximal value for the causal effect and call this value γ^* .⁶ In the spirit of our previous prior which orders the direct and indirect effects, we could adopt a prior of the form:

$$(4) \quad \alpha_1 | \beta_1 \sim \mathcal{TN}_{(-|\gamma^*\beta_1|, |\gamma^*\beta_1|)}(0, \underline{V}_{\alpha_1}).$$

This specification may help in terms of prior elicitation, as the product $\gamma\beta_1$ is a marginal effect directly comparable in scale to α_1 and researchers, either from past work or from informed prior beliefs, are likely to have a reasonable sense about values of the causal effect γ . If it is desired to have some degree of uncertainty in the selection of γ^* , a hierarchical specification could also be adopted, as in

$$(5) \quad \alpha_1 | \beta_1, \gamma^* \sim \mathcal{TN}_{(-|\gamma^*\beta_1|, |\gamma^*\beta_1|)}(0, \underline{V}_{\alpha_1}), \quad \gamma^* \sim \mathcal{TN}_{(\underline{\gamma}, \bar{\gamma})}(\mu_{\gamma^*}, V_{\gamma^*}).$$

The priors in (4) and (5) do not exclude 0 from the support of the γ posterior distribution, although they no longer enforce the ordering of direct and indirect effects. They instead place bounds on α_1 , which are revised by the data through learning about β_1 , by thinking about values the primary “indirect” marginal effect is likely to take. The prior in (4) is quite similar to that used in the work of Conley et al (2012), who consider priors of the forms $\alpha_1 \sim N(\mu, \delta^2)$ and $\alpha_1 | \gamma \sim N(0, \delta^2 \gamma^2)$. This second version of their prior, by choosing δ to be small, embodies the idea that the direct effect of z on y is likely to be smaller than the causal effect γ of s on y . In practice one can, of course, end up with approximately the same priors through appropriate choices of δ and γ^* . We believe that representing the prior as we have in (4) and (5) could be useful to the practitioner for elicitation purposes, as one might choose γ^* , the approximate counterpart of δ in the Conley et al prior, as a perceived large or maximal value of the causal effect — a parameter for which the researcher may have reasonable prior information. In what follows,

⁵This point will be demonstrated formally in section 3.1.

⁶Given this interpretation, the marginal prior for γ should then be chosen to be consistent with the selection of γ^* .

we will employ all of these conditional priors in (3) - (5) our empirical work and investigate how posterior results change under different sets of prior information. We will primarily focus on (3) in the exposition, although this is essentially without loss of generality.

In addition to these alternate priors over α_1 , we also consider the use of a potentially more informative prior for the error correlation ρ_{uv} . Specifically, instead of the usual conditionally conjugate Wishart analysis for the error covariance matrix, we also consider truncated normal priors for this parameter:

$$(6) \quad \rho_{uv} \sim TN_{(\underline{\rho}_l, \underline{\rho}_u)}(\underline{\mu}_\rho, \underline{V}_\rho),$$

with $\underline{\rho}_l$ and $\underline{\rho}_u$ denoting the lower and upper truncation limits, respectively.

Motivation for the prior in (6) comes both from our own personal experience (which suggests that, when the model is sufficiently specified to enable estimation of ρ_{uv} , the estimated value is typically “small”) as well as related approaches in the literature.

To make this connection to the literature more specific, consider that, as an alternative to traditional IV / 2SLS, a comparatively small set of studies have pursued the “kitchen sink” approach to identification, estimation and inference. In this approach, a rich set of controls are included in \mathbf{x} and, conditioned on this extensive set of controls, any remaining correlation between u and ϵ is believed to be quite small. As a result, one may approximately assume that $\rho_{uv} = 0$, thus escaping the need for instruments, and pursuing estimation of (1) using traditional methods such as OLS. Deaerden et al. (2002) is a prominent, well-crafted example⁷ that exemplifies this identification strategy.

We recognize that such an approach simply achieves identification through different prior information: instead of searching for a z for which $\alpha_1 = 0$ can credibly be imposed, the researcher instead makes the case that \mathbf{x} is rich enough so that $\rho_{uv} = 0$. Although this identification strategy seems to remain the exception to the rule, it does highlight the idea that $\rho_{uv} = 0$ is an alternate approach to identification, and in many cases, this correlation is likely to be modest in value. Given that the set of controls available in \mathbf{x} is often quite extensive, we therefore consider priors that allow for the imposition of this type of information, and will investigate the nature of learning about γ and α_1 when such information is imposed. Of course, sensitivity analysis can and should be conducted to see how changes in this type of belief influences posterior results. Finally, for the remaining parameters, we will specify normal priors for the elements of γ , $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, and inverse

⁷They write (page 3), “In our view, the best way to deal with endogeneity issues with data such as ours is to control for the variables that are likely to be driving [school] selection...”

gamma priors for the variance parameters σ_u^2 and σ_v^2 . These will be described in full detail in the following section.

2.2. Relation to the Existing Literature. We begin by recognizing that there are a number of applied Bayesian studies in the treatment-response literature that are relevant to our work, and share similarities with our MCMC implementation, including Li (1998), Chib and Hamilton (2000), (2004), Li and Poirier (2003a) (2003b), Geweke et. al (2003), Munkin and Trivedi (2003) (2008), Deb et. al (2006), and Kline and Tobias (2008), among others. These studies, however, do not deal with issues of imperfect instruments, which is the central theme of this paper.

Our idea to allow for some “imperfectness” in the instrument is not new, and indeed, similar methodologies have been recently advanced by other authors. Koop and Poirier (1997) and Poirier (1998), for example, discuss Bayesian learning in partially identified models, the latter study specifically addressing learning in the related generalized Roy model of selectivity. Ashley (2009) derives asymptotic results for the IV estimator under different assumptions regarding the correlation structure between the outcome error and instruments. Numerous non-Bayesian papers employ (typically nonparametric) bounding strategies and illustrate how learning regarding the structural parameter of interest takes place under different prior assumptions, such as monotonicity [e.g., Manski and Pepper (2000) and Gundersen and Kreider (2009), among others]. Nevo and Rosen (2010) derive bounds on the causal effect parameter given sign restrictions on the correlations between the endogenous variable and the error and the instrument and the error, and potentially, an ordering of the magnitudes of these two correlations.

As mentioned previously, Conley, Hansen and Rossi (2012) employs methods that are quite similar to our own and one of the four approaches they discuss to model instrument imperfection is, in fact, a fully Bayesian analysis. In an interesting article Kraay (2012) also considers a Bayesian analysis of a just-identified version of the continuous outcome, continuous endogenous variable model like the one we consider here. Just identification produces a one-to-one mapping between the reduced form and structural coefficients, and thus enables the researcher to work directly with (and derive analytical results for) the reduced form posteriors. These results can then be used to back out implied marginal posterior distributions for structural coefficients of interest. Kraay (2012) documents a rather considerable loss in the precision of marginal posterior distributions as one departs from point identification, and also describes how such losses are related to quantities like the sample size and the strength of the instrument.

The primary focus of our paper is not in crafting priors to allow for degrees of instrument imperfection or to theoretically document changes in posterior distributions as one departs from point identification. We do not require just identification, we allow for the specification of proper priors directly over the structural rather than reduced form parameters, and we also consider priors that preclude fully analytic derivations. As such, we take up cases where MCMC is used to conduct the model fitting, similar to the situation considered by Conley et al (2012). Our paper, with a slightly different focus than these important papers on which we build, thus emphasizes computational considerations and seeks to provide an improved method for posterior calculation given nothing more than the “standard” MCMC output.

In the following section we formally discuss the mechanism for learning about non-identified “structural” parameters of interest like the causal effect γ in this partially identified model. This exercise also reveals a quite useful computational strategy for calculating various marginal posterior distributions. This procedure will combine analytic results associated with the forms of the conditional priors with available well-mixing posterior simulations associated with the *identifiable* model parameters. We find and demonstrate that posterior calculations based solely on simulations from the non-identified model can be quite unstable and can produce unacceptably large numerical standard errors. We discuss an efficient alternative to this problem in the following section. We believe that doing so will provide researchers with an efficient recipe for calculating posterior quantities of interest when instruments are allowed to be imperfect, and thus allow researchers to reliably report a range of posterior results under different types of prior information.

3. LEARNING AND POSTERIOR CALCULATION

Recall that the parameters of interest are $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \boldsymbol{\alpha}_1, \beta_0, \beta_1, \boldsymbol{\beta}_2, \gamma, \sigma_u^2, \sigma_v^2, \rho_{uv})'$, while the identified parameters are $\boldsymbol{\psi} = (a_0, a_1, \boldsymbol{a}_2, \beta_0, \beta_1, \boldsymbol{\beta}_2, b, \sigma_v^2, c^2)'$. What we are primarily interested in are marginal posterior densities of the forms $p(\theta_i | \mathbf{y}, \mathbf{s})$ and characterization of the marginal posterior for the causal effect parameter $[p(\gamma | \mathbf{y}, \mathbf{s})]$ in particular. For $\theta_i = \beta_0, \beta_1, \boldsymbol{\beta}_2$ or σ_v^2 , the marginal posterior densities $p(\theta_i | \mathbf{y}, \mathbf{s})$ and their features can be calculated in the usual way from the simulated output, as those components of $\boldsymbol{\theta}$ are identifiable. For elements in $\boldsymbol{\theta}$ that are not identifiable, we seek to investigate their marginal posteriors in greater detail. To this end, we

note that, for a non-identified θ_i :

$$\begin{aligned}
 p(\theta_i | \mathbf{y}, \mathbf{s}) &= \int p(\theta_i, \boldsymbol{\psi} | \mathbf{y}, \mathbf{s}) d\boldsymbol{\psi} \\
 (7) \qquad \qquad \qquad &= \int p(\theta_i | \boldsymbol{\psi}, \mathbf{y}, \mathbf{s}) p(\boldsymbol{\psi} | \mathbf{y}, \mathbf{s}) d\boldsymbol{\psi}.
 \end{aligned}$$

In the last equation, the rightmost density is simply the posterior for the identifiable parameters. The leftmost density in the integrand, $p(\theta_i | \boldsymbol{\psi}, \mathbf{y}, \mathbf{s})$, with a little work, can be shown to equal the conditional prior $p(\theta_i | \boldsymbol{\psi})$. To see this, note:

$$\begin{aligned}
 (8) \qquad \qquad \qquad p(\theta_i | \boldsymbol{\psi}, \mathbf{y}, \mathbf{s}) &= \frac{p(\mathbf{y}, \mathbf{s} | \theta_i, \boldsymbol{\psi}) p(\boldsymbol{\psi}, \theta_i)}{p(\boldsymbol{\psi}, \mathbf{y}, \mathbf{s})} \\
 &= \frac{p(\mathbf{y}, \mathbf{s} | \boldsymbol{\psi}) p(\boldsymbol{\psi}, \theta_i)}{p(\mathbf{y}, \mathbf{s} | \boldsymbol{\psi}) p(\boldsymbol{\psi})} \\
 &= \frac{p(\boldsymbol{\psi}, \theta_i)}{p(\boldsymbol{\psi})} \\
 &= p(\theta_i | \boldsymbol{\psi}),
 \end{aligned}$$

where the second step follows since the data density does not depend on θ given $\boldsymbol{\psi}$. Substituting this into (7), we obtain:

$$(9) \qquad \qquad \qquad p(\theta_i | \mathbf{y}, \mathbf{s}) = \int p(\theta_i | \boldsymbol{\psi}) p(\boldsymbol{\psi} | \mathbf{y}, \mathbf{s}) d\boldsymbol{\psi},$$

which simply shows that the posterior density of interest is the posterior expectation of the conditional prior. The result in (9) echoes the general conclusions of Poirier (1998) and Moon and Schorfheide (2011), who argue that, *conditionally*, no updating of the structural parameter of interest will take place. Moon and Schorfheide (2011) also demonstrate that, asymptotically, the marginal posterior for θ_i in (9) will equal the conditional prior evaluated at the MLE.

3.1. Issues Surrounding Posterior Computation. Objects of interest like (9) can be calculated in a variety of ways. First, and most naturally, one can simply fit the model in the $\boldsymbol{\theta}$ -parameterization using Gibbs sampling or, if priors like those in equations (3)-(6) are employed, Gibbs/ Metropolis Hastings, and directly use the post-convergence simulations for the θ_i to calculate posterior statistics of interest. We suspect that this would be the prevailing approach in the applied literature.

Although this approach is quite natural, a problem with its use is that simulations of the non-identified elements of θ_i often prove to be highly autocorrelated, rendering such “direct” estimates of posterior means and standard deviations unreliable. Furthermore, the severity of

autocorrelation among the non-identified simulations increases with the degree of uncertainty surrounding the validity of the instruments. We will document this issue in the following section.

While posterior simulations for the non-identified components of $\boldsymbol{\theta}$ can mix very poorly, the identified parameters $\boldsymbol{\psi}$ usually enjoy very good mixing properties. This suggests the potential for significant improvement in performance provided the *identified* simulations can somehow be used for posterior computation purposes in place of the non-identified simulations.

We take advantage of this insight and introduce a second method for calculating posterior statistics and densities of the non-identified elements of $\boldsymbol{\theta}$. Equation (9) reveals that the posterior simulations associated with the *identified* parameter vector $\boldsymbol{\psi}$ can be used to calculate the desired marginal posterior distribution, provided the conditional prior $p(\theta_i|\boldsymbol{\psi})$ is known or easily sampled. In what follows, then, we describe a scheme for calculating this conditional prior and thus offer a “semi-analytic” scheme for calculating the marginal posterior $p(\theta_i|\mathbf{y}, \mathbf{s})$. Specifically, we will compute the one-dimensional density $p(\theta_i|\boldsymbol{\psi})$ on a grid. From (9), we can then average these densities over a sufficiently large number of posterior draws from $p(\boldsymbol{\psi}|\mathbf{y}, \mathbf{s})$ to approximate $p(\theta_i|\mathbf{y}, \mathbf{s})$ arbitrarily well.

To operationalize this semi-analytic strategy, we first need to characterize the conditional prior $p(\theta_i|\boldsymbol{\psi})$. To this end, suppose the prior $p(\boldsymbol{\theta})$ for the structural parameters is given and, to fix ideas, suppose we wish to derive the marginal posterior distribution of the causal effect parameter $p(\gamma|\mathbf{y}, \mathbf{s})$. In this regard, let us consider the transformation $\mathbf{g}_\gamma : \boldsymbol{\theta} \mapsto \boldsymbol{\psi}_\gamma$:

$$\begin{pmatrix} \gamma \\ \sigma_v^2 \\ \beta_0 \\ \beta_1 \\ \boldsymbol{\beta}_2 \\ \alpha_0 \\ \alpha_1 \\ \boldsymbol{\alpha}_2 \\ \rho_{uv} \\ \sigma_u^2 \end{pmatrix} \mapsto \begin{pmatrix} \gamma \\ \sigma_v^2 \\ \beta_0 \\ \beta_1 \\ \boldsymbol{\beta}_2 \\ g_{a_0}(\boldsymbol{\theta}) \\ g_{a_1}(\boldsymbol{\theta}) \\ g_{a_2}(\boldsymbol{\theta}) \\ g_b(\boldsymbol{\theta}) \\ g_{c^2}(\boldsymbol{\theta}) \end{pmatrix} \equiv \boldsymbol{\psi}_\gamma,$$

where

$$g_{a_0}(\boldsymbol{\theta}) = \alpha_0 - \frac{\rho_{uv}\sigma_u}{\sigma_v}\beta_0, \quad g_{a_1}(\boldsymbol{\theta}) = \alpha_1 - \frac{\rho_{uv}\sigma_u}{\sigma_v}\beta_1, \quad g_{\mathbf{a}_2}(\boldsymbol{\theta}) = \boldsymbol{\alpha}_2 - \frac{\rho_{uv}\sigma_u}{\sigma_v}\boldsymbol{\beta}_2,$$

$$g_b(\boldsymbol{\theta}) = \gamma + \frac{\rho_{uv}\sigma_u}{\sigma_v}, \quad \text{and} \quad g_{c^2}(\boldsymbol{\theta}) = (1 - \rho_{uv}^2)\sigma_u^2.$$

It is easy to show that the inverse of the transformation $\mathbf{g}_\gamma^{-1} : \boldsymbol{\psi}_\gamma \mapsto \boldsymbol{\theta}$ is given by

$$\begin{pmatrix} \gamma \\ \sigma_v^2 \\ \beta_0 \\ \beta_1 \\ \boldsymbol{\beta}_2 \\ a_0 \\ a_1 \\ \mathbf{a}_2 \\ b \\ c^2 \end{pmatrix} \mapsto \begin{pmatrix} \gamma \\ \sigma_v^2 \\ \beta_0 \\ \beta_1 \\ \boldsymbol{\beta}_2 \\ g_{\alpha_0}^{-1}(\boldsymbol{\psi}_\gamma) \\ g_{\alpha_1}^{-1}(\boldsymbol{\psi}_\gamma) \\ g_{\boldsymbol{\alpha}_2}^{-1}(\boldsymbol{\psi}_\gamma) \\ g_{\rho_{uv}}^{-1}(\boldsymbol{\psi}_\gamma) \\ g_{\sigma_u^2}^{-1}(\boldsymbol{\psi}_\gamma) \end{pmatrix},$$

where

$$\alpha_0 = g_{\alpha_0}^{-1}(\boldsymbol{\psi}_\gamma) = a_0 + (b - \gamma)\beta_0,$$

$$\alpha_1 = g_{\alpha_1}^{-1}(\boldsymbol{\psi}_\gamma) = a_1 + (b - \gamma)\beta_1,$$

$$\boldsymbol{\alpha}_2 = g_{\boldsymbol{\alpha}_2}^{-1}(\boldsymbol{\psi}_\gamma) = \mathbf{a}_2 + (b - \gamma)\boldsymbol{\beta}_2$$

$$\rho_{uv} = g_{\rho_{uv}}^{-1}(\boldsymbol{\psi}_\gamma) = \frac{(b - \gamma)\sigma_v}{\sqrt{c^2 + (b - \gamma)^2\sigma_v^2}},$$

$$\sigma_u^2 = g_{\sigma_u^2}^{-1}(\boldsymbol{\psi}_\gamma) = c^2 + (b - \gamma)^2\sigma_v^2.$$

Let $\mathbf{J}(\boldsymbol{\psi}_\gamma)$ be the Jacobian matrix associated with the transformation from $\boldsymbol{\theta}$ to $\boldsymbol{\psi}_\gamma$:

$$\mathbf{J}(\boldsymbol{\psi}_\gamma) = \left(\frac{\partial \theta_i}{\partial \psi_{\gamma,j}} \right) = \left(\begin{array}{c|c} \mathbf{J}_{\boldsymbol{\psi}_\gamma}^{11} & \mathbf{J}_{\boldsymbol{\psi}_\gamma}^{12} \\ \mathbf{J}_{\boldsymbol{\psi}_\gamma}^{21} & \mathbf{J}_{\boldsymbol{\psi}_\gamma}^{22} \end{array} \right)$$

$$= \left(\begin{array}{cccc|cccccc} 1 & 0 & 0 & 0 & \mathbf{0}_{1 \times k_x} & 0 & 0 & \mathbf{0}_{1 \times k_x} & 0 & 0 \\ 0 & 1 & 0 & 0 & \mathbf{0}_{1 \times k_x} & 0 & 0 & \mathbf{0}_{1 \times k_x} & 0 & 0 \\ 0 & 0 & 1 & 0 & \mathbf{0}_{1 \times k_x} & 0 & 0 & \mathbf{0}_{1 \times k_x} & 0 & 0 \\ 0 & 0 & 0 & 1 & \mathbf{0}_{1 \times k_x} & 0 & 0 & \mathbf{0}_{1 \times k_x} & 0 & 0 \\ \mathbf{0}_{k_x \times 1} & \mathbf{0}_{k_x \times 1} & \mathbf{0}_{k_x \times 1} & \mathbf{0}_{k_x \times 1} & \mathbf{I}_{k_x} & \mathbf{0}_{k_x \times 1} & \mathbf{0}_{k_x \times 1} & \mathbf{0}_{k_x \times k_x} & \mathbf{0}_{k_x \times 1} & \mathbf{0}_{k_x \times 1} \\ \hline -\beta_0 & 0 & -\gamma & 0 & \mathbf{0}_{1 \times k_x} & 1 & 0 & \mathbf{0}_{1 \times k_x} & \beta_0 & 0 \\ -\beta_1 & 0 & 0 & -\gamma & \mathbf{0}_{1 \times k_x} & 0 & 1 & \mathbf{0}_{1 \times k_x} & \beta_1 & 0 \\ \beta_2 & \mathbf{0}_{k_x \times 1} & \mathbf{0}_{k_x \times 1} & \mathbf{0}_{k_x \times 1} & \gamma \mathbf{I}_{k_x} & \mathbf{0}_{k_x \times 1} & \mathbf{0}_{k_x \times 1} & \mathbf{I}_{k_x} & \beta_2 & \mathbf{0}_{k_x \times 1} \\ \frac{\partial \rho_{uv}}{\partial \gamma} & \frac{\partial \rho_{uv}}{\partial \sigma_u^2} & 0 & 0 & \mathbf{0}_{1 \times k_x} & 0 & 0 & \mathbf{0}_{1 \times k_x} & \frac{\partial \rho_{uv}}{\partial b} & \frac{\partial \rho_{uv}}{\partial c^2} \\ -2(b-\gamma)\sigma_v^2 & (b-\gamma)^2 & 0 & 0 & \mathbf{0}_{1 \times k_x} & 0 & 0 & \mathbf{0}_{1 \times k_x} & \frac{\partial \sigma_u^2}{\partial b} & 1 \end{array} \right).$$

Since

$$\det[\mathbf{J}(\boldsymbol{\psi}_\gamma)] = \det[\mathbf{J}_{\boldsymbol{\psi}_\gamma}^{11}] \det[\mathbf{J}_{\boldsymbol{\psi}_\gamma}^{22} - \mathbf{J}_{\boldsymbol{\psi}_\gamma}^{21} (\mathbf{J}_{\boldsymbol{\psi}_\gamma}^{11})^{-1} \mathbf{J}_{\boldsymbol{\psi}_\gamma}^{12}] = \det[\mathbf{J}_{\boldsymbol{\psi}_\gamma}^{22}],$$

and applying this result again to the lower (2,2) block, it follows that the Jacobian term is surprisingly simple:

$$|\det \mathbf{J}(\boldsymbol{\psi}_\gamma)| = \left| \frac{\partial \rho_{uv}}{\partial b} - \frac{\partial \rho_{uv}}{\partial c^2} \frac{\partial \sigma_u^2}{\partial b} \right|$$

$$= \frac{\sigma_v}{\sqrt{c^2 + (b-\gamma)^2 \sigma_v^2}}.$$

With the Jacobian and inverse transformations in place, the conditional prior density can be calculated up to proportionality:

$$(10) \quad p(\gamma | \boldsymbol{\psi}) \propto p(\gamma, \boldsymbol{\psi}) = p(\boldsymbol{\psi}_\gamma) = |\det \mathbf{J}(\boldsymbol{\psi}_\gamma)| p(\mathbf{g}_\gamma^{-1}(\boldsymbol{\psi}_\gamma)),$$

where all terms on the rightmost side of this equation can be evaluated given $\boldsymbol{\psi}_\gamma = (\boldsymbol{\psi}, \gamma)$.

Equation (10) provides the foundation for a general procedure for calculating marginal posteriors and their associated statistics. The specific form of (10), of course, depends on the form of the prior $p(\boldsymbol{\theta})$. Below, we consider results based upon the prior in (3) as well as widely-employed

priors for the remaining model parameters:

$$\begin{aligned}
\gamma &\sim N(\underline{\mu}_\gamma, \underline{V}_\gamma) \\
\sigma_v^2 &\sim IG(\underline{s}_{1v}, \underline{s}_{2v}) \\
\sigma_u^2 &\sim IG(\underline{s}_{1u}, \underline{s}_{2u}) \\
\boldsymbol{\beta} \equiv [\beta_0 \ \beta_1 \ \boldsymbol{\beta}'_2]' &\sim N(\underline{\mu}_\beta, \underline{V}_\beta) \\
\alpha_0 &\sim N(\underline{\mu}_{\alpha_0}, \underline{V}_{\alpha_0}) \\
\boldsymbol{\alpha}_2 &\sim N(\underline{\mu}_{\alpha_2}, \underline{V}_{\alpha_2}) \\
\rho_{uv} &\sim TN_{(\rho_l, \rho_u)}(\underline{\mu}_\rho, \underline{V}_\rho) \\
\alpha_1 | \gamma, \beta_1 &\sim TN_{(-|\gamma\beta_1|, |\gamma\beta_1|)}(\underline{\mu}_{\alpha_1}, \underline{V}_{\alpha_1}).
\end{aligned}$$

From (10), we are then able to obtain our conditional prior up to proportionality:⁸

$$\begin{aligned}
p(\gamma | \boldsymbol{\psi}) &\propto \left(\frac{\sigma_v}{\sqrt{c^2 + (b - \gamma)^2 \sigma_v^2}} \right) [c^2 + (b - \gamma)^2 \sigma_v^2]^{-(s_{1u} + 1)} \left[\Phi \left(\frac{|\gamma\beta_1| - \underline{\mu}_{\alpha_1}}{\sqrt{\underline{V}_{\alpha_1}}} \right) - \Phi \left(\frac{-|\gamma\beta_1| - \underline{\mu}_{\alpha_1}}{\sqrt{\underline{V}_{\alpha_1}}} \right) \right]^{-1} \\
&\exp \left(-\frac{1}{2} \left(\left[\frac{\gamma - \underline{\mu}_\gamma}{\sqrt{\underline{V}_\gamma}} \right]^2 + \left[\frac{a_0 + (b - \gamma)\beta_0 - \underline{\mu}_{\alpha_0}}{\sqrt{\underline{V}_{\alpha_0}}} \right]^2 + \left[\frac{a_1 + (b - \gamma)\beta_1 - \underline{\mu}_{\alpha_1}}{\sqrt{\underline{V}_{\alpha_1}}} \right]^2 + \right. \right. \\
&2[\underline{s}_{2u}^{-1}(c^2 + (b - \gamma)^2 \sigma_v^2)]^{-1} + [\mathbf{a}_2 + (b - \gamma)\boldsymbol{\beta}_2 - \underline{\mu}_{\alpha_2}]' \underline{V}_{\alpha_2}^{-1} [\mathbf{a}_2 + (b - \gamma)\boldsymbol{\beta}_2 - \underline{\mu}_{\alpha_2}] + \\
&\left. \left. \left[\frac{(b - \gamma)\sigma_v}{\sqrt{\underline{V}_\rho} [c^2 + (b - \gamma)^2 \sigma_v^2]} - \frac{\underline{\mu}_\rho}{\sqrt{\underline{V}_\rho}} \right]^2 \right) \right) I(|a_1 + (b - \gamma)\beta_1| < |\gamma\beta_1|) \\
&I \left(\rho_l \leq \frac{(b - \gamma)\sigma_v}{\sqrt{c^2 + (b - \gamma)^2 \sigma_v^2}} \leq \rho_u \right),
\end{aligned}$$

where, consistent with (3), we anticipate choosing $\underline{\mu}_{\alpha_1} = 0$.

Of course, other choices for $p(\boldsymbol{\theta})$ will lead to different conditional priors $p(\gamma | \boldsymbol{\psi})$, although the general procedure for its calculation will remain the same: For a given $\boldsymbol{\psi} \sim p(\boldsymbol{\psi} | \mathbf{y}, \mathbf{s})$, we calculate $p(\gamma | \boldsymbol{\psi})$ via a change of variables up to proportionality over a fine, discrete grid of γ values. The resulting collection of ordinates is then normalized to obtain a proper discrete conditional prior

⁸This can be simplified further; we present it in this way as it is the direct result of applying (10), given the Jacobian and inverse mappings.

density. Following (9), this process is repeated for all $\boldsymbol{\psi}$ post-convergence simulations, producing a collection of conditional prior densities. These are then averaged to numerically approximate the marginal posterior. Posterior means, standard deviations and other statistics can also be easily calculated from this approach.

Calculations for the remaining non-identified elements of $\boldsymbol{\theta}$ other than the causal effect γ follow a similar procedure to what we have just described. Each new parameter $\theta_i \neq \gamma$ requires us to go through a similar change of variables exercise, and we provide these details for all non-identified parameters in the appendix.⁹ Similar derivations can also be performed when priors like those in (4) or (5) are employed; we do not present these here in the interest of brevity, although such results are available upon request.

3.2. Prior Sensitivity and Other Computational Possibilities. We close with a few observations regarding this method. First, the sensitivity of posterior results to changes in the prior can be *approximately* investigated in sufficiently large samples by simply changing the hyperparameters in $p(\theta_i|\boldsymbol{\psi})$ and repeating the previous process. That is, if the sample size is large enough, and the implied priors on $\boldsymbol{\psi}$ are sufficiently dispersed, we would not expect $p(\boldsymbol{\psi}|\mathbf{y}, \mathbf{s})$ to change significantly under the new prior, whereas $p(\theta_i|\boldsymbol{\psi})$ will certainly be affected. So, at least as a quick approximation, the researcher could assess the influence of the prior on posterior results without having to refit the model under each new prior under consideration, and can instead simply follow the procedure just outlined with the same set of $\boldsymbol{\psi}$ values but with updated definitions of the prior hyperparameters. This is decidedly not the case if one simply fits the model in the $\boldsymbol{\theta}$ -parameterization and uses the θ_i simulations to calculate quantiles of interest; in this case, the model should be refit under each new prior, or some alternate approach must be used to investigate prior sensitivity.

Second, there seems to be the potential for a variety of “intermediate” solutions which neither use the $\boldsymbol{\theta}$ simulations directly nor need to analytically characterize the conditional prior.¹⁰ To see this, consider calculation of

$$(11) \quad E(\theta_i|\mathbf{y}, \mathbf{s}) = E_{\boldsymbol{\psi}|\mathbf{y}, \mathbf{s}} [E(\theta_i|\boldsymbol{\psi})].$$

⁹A reparameterization is required in order to apply this technique to calculate statistics associated with the variance parameter σ_u^2 . These details are contained in the appendix.

¹⁰We can imagine, in more elaborate models, where it will be quite difficult to determine the Jacobian term. In these cases, the semi-analytic methods like we have described here would be difficult to apply.

Upon fitting the model, a set of posterior simulations for $\boldsymbol{\psi}$ are produced. Numerically calculating (11) based on these simulations then requires evaluation of the conditional prior expectation. One could approximate the conditional expectation $E(\theta_i|\boldsymbol{\psi} = \boldsymbol{\psi}^{(m)})$ a number of ways, perhaps using kernel methods, given a set of simulations from the joint prior. Specifically, note $p(\boldsymbol{\theta}, \boldsymbol{\psi}) = p(\boldsymbol{\psi}|\boldsymbol{\theta})p(\boldsymbol{\theta}) = \mathbb{1}[\boldsymbol{\psi} = \mathbf{g}(\boldsymbol{\theta})]p(\boldsymbol{\theta})$. Thus, a draw from the *joint prior* can be obtained by first sampling $\boldsymbol{\theta}^{(j)}$ from $p(\boldsymbol{\theta})$ and then calculating $\boldsymbol{\psi}^{(j)} = \mathbf{g}(\boldsymbol{\theta}^{(j)})$. Given (a very large) J such simulations, we can calculate:

$$\widehat{\mu}_m = E(\theta_i|\widehat{\boldsymbol{\psi}} = \boldsymbol{\psi}^{(m)}) = \frac{\sum_{j=1}^J \theta_i^{(j)} K \left[(\boldsymbol{\psi}^{(j)} - \boldsymbol{\psi}^{(m)})' \widehat{\Sigma}^{-1} (\boldsymbol{\psi}^{(j)} - \boldsymbol{\psi}^{(m)}) \right]}{\sum_{j=1}^J K \left[(\boldsymbol{\psi}^{(j)} - \boldsymbol{\psi}^{(m)})' \widehat{\Sigma}^{-1} (\boldsymbol{\psi}^{(j)} - \boldsymbol{\psi}^{(m)}) \right]}$$

for a given $\boldsymbol{\psi}^{(m)} \sim p(\boldsymbol{\psi}|\mathbf{y}, \mathbf{s})$. In the above, $\boldsymbol{\psi}^{(m)}$ denotes the m^{th} post-convergence simulation of the identified parameter vector, K denotes a kernel function and $\widehat{\Sigma}$ denotes an estimate of the posterior covariance matrix of $\boldsymbol{\psi}$. The posterior mean can then be estimated as:

$$E(\widehat{\theta}_i|\widehat{\mathbf{y}}, \mathbf{s}) = \frac{1}{M} \sum_{m=1}^M \widehat{\mu}_m.$$

This approach has the potential to be advantageous in that it requires very little in terms of inputs: a set of simulations of the identified parameters $\boldsymbol{\psi}$ from the joint posterior and a series of draws from the joint prior. These are easily obtained. In addition, it exploits the result that the conditional posterior is simply the conditional prior, and uses (iid) simulations from the prior while avoiding use of the poorly mixing $\boldsymbol{\theta}$ posterior simulations in the calculation. In high dimensional situations, however, the usual dimensionality concerns will apply and will mitigate the potential benefits of this method. We do not explore the possibility of using this approach in the present paper, given that we were able to derive the conditional prior analytically. In the following section, we illustrate the efficiency gains associated with calculating posterior moments using our semi-analytic method, while documenting the potential for considerable inaccuracy when directly making use of the $\boldsymbol{\theta}$ posterior draws.

4. ILLUSTRATIVE APPLICATION

In our illustrative application we reconsider the results of Kline and Tobias (2008). These authors were interested in estimating the causal effect of BMI (Body Mass Index) on earnings, a question which has received considerable attention in the labor economics literature [e.g., Cawley (2004)].

Kline and Tobias (2008), using data from the British Cohort Study, employed maternal and paternal BMI as instruments for the respondent’s BMI in the log wage outcome equation. The argument behind their choice was that parental BMI surely correlates with child BMI, yet conditioned upon the BMI of the employee (child), parental BMI plays no direct role in his / her earnings production. While there has been a precedent for such instruments in the literature (as Cawley (2004), for example, employs sibling BMI as an instrument), and a variety of other controls were employed to bolster the case for the validity of the instruments, it remains entirely possible that unobserved family background characteristics may simultaneously correlate with child BMI and child wages, undermining the validity of the instrumental variable strategy. With this possibility in mind, we reanalyze their data for a subsample of females, which contained $n = 1,782$ observations.

Before discussing posterior results from this data, and then documenting sensitivity of those results to the prior, we first illustrate differences in the mixing properties of the identified and non-identified posterior simulations. These results are obtained upon using priors like those in Conley et al. which specify $\alpha_{1i} \stackrel{iid}{\sim} N(0, .05^2)$, $i = 1, 2$.¹¹ We thus begin by choosing priors for the “direct” parental BMI effects that are quite spread out, though regarding it very unlikely that a one unit increase in parental BMI could raise or lower hourly wages by, say, more than 10 percent. We choose this rather dispersed prior as a starting point to starkly illustrate the potential for estimation inaccuracy when using standard Gibbs with such weak prior information. Priors for the remaining parameters of the model are less critical, but for the intercepts of each equation we specify $\alpha_0, \beta_0 \sim N(0, 5)$, all other remaining slope coefficients are assigned independent $N(0, 1)$ priors and $\Sigma \sim IW(4, \mathbf{I}_2)$. Unless otherwise noted, the priors just listed are used in all of the calculations contained in this section.

[Figure 1 about here.]

[Figure 2 about here.]

Figure 1 presents lagged autocorrelations for a selection of non-identified parameters. These include the causal effect parameter γ , the error correlation ρ_{uv} as well as the coefficients on the instrumental variables α_{11} and α_{12} . As the figure clearly reveals, the mixing of these simulations is very poor: even if we were to use every 1,000th iteration from the sampler, the autocorrelation between those draws would be in excess of .8. This suggests that simply fitting the model in the θ parameterization, and directly using the simulations that are produced from the sampler will yield

¹¹In this application we have two potential instruments (both maternal and paternal BMI).

very unreliable estimates of posterior moments. This is particularly true given conventions that are sometimes followed in applied work, where 50,000 or 100,000 simulations are often believed to be sufficient to safeguard the researcher from most simulation-related problems. This is decidedly not the case in this particular model once we depart from point identification.¹²

In contrast to Figure 1, Figure 2 plots lagged autocorrelations for four *identifiable* parameters: a_1 , a_2 , b and c^2 .¹³As can be seen from this figure, the lag-1 autocorrelations are essentially zero and mimic what we would observe in an *iid* sample. This figure reveals great promise for our proposed approach to posterior calculation: the identifiable parameters mix essentially as an iid sample, and these are the only simulations that we will use to calculate posterior moments for the non-identified parameters. In fact, in the spirit of Rao-Blackwell (since we use the form of the conditional prior to calculate the marginal posterior), we expect that estimates obtained from our method will be as good as, if not better than, those we would obtain from an *iid* sample of equal size.

Table 1 presents results of an experiment of sorts using the actual Kline and Tobias data. We perform this experiment as follows. We first fit the model using a standard two-block Gibbs sampler¹⁴ and obtain one million post-convergence simulations. We then use these simulations to calculate values of the identified quantities ψ for each post-convergence iteration. Finally, we employ our semi-analytic method to calculate posterior moments of all the parameters and treat these estimates as if they were the exact posterior means.

Given these posterior moments, we ask the following question: how many post-convergence simulations would be required in order to achieve a NSE that is less than one percent of that mean? We ask this question of both our semi-analytic method as well as the “standard” Gibbs method that uses the θ_i simulations directly when calculating the first moment. The choice of one percent is a rather exacting standard, but seems reasonable given that posterior moment estimation uncertainty is often ignored in applied work (and thus one would hope that the estimated mean

¹²Kraay (2012) makes an important related point, noting that posterior standard deviations for parameters like the causal effect are strongly affected by even moderate degrees of prior uncertainty regarding the validity of the instruments. This increase, of course, also contributes to an increase in NSE’s for estimated moments.

¹³Recall, from section 2, their definitions: $a_1 \equiv \alpha_1 - \frac{\rho_{uv}\sigma_u}{\sigma_v}\beta_1$, $\mathbf{a}_2 \equiv \boldsymbol{\alpha}_2 - \frac{\rho_{uv}\sigma_u}{\sigma_v}\boldsymbol{\beta}_2$, $b \equiv \gamma + \frac{\rho_{uv}\sigma_u}{\sigma_v}$, $c^2 \equiv (1 - \rho_{uv}^2)\sigma_u^2$.

¹⁴We can use standard Gibbs when a prior like $\alpha_1 \sim N(0, V_\alpha)$ is employed; otherwise we employ a Gibbs / Metropolis-Hastings scheme for posterior simulation. Details associated with model fitting are not central to this paper, but are available upon request.

is quite close to the actual posterior mean), and simply offers a reasonable metric that can be used for comparison purposes.

As shown in Table 1, for example, we would require approximately 3.4 *billion* simulations to achieve this level of precision for the causal effect parameter γ using the Gibbs output directly! Our method, however, can achieve this level of precision in a manageable 146,000 simulations. Coefficients on the endogenous variable BMI, the potential instruments (MomBMI, DadBMI) and the error correlation ρ_{uv} are, not surprisingly, the most problematic of the non-identified parameters in terms of accurately calculating their posterior means.

The fourth column of the table also offers a measure of the computation time required to achieve this level of numerical accuracy. These times are calculated when assuming that it takes 1 second to obtain 100 post-convergence simulations in the Gibbs sampler¹⁵ and then determining how long it would take to achieve this level of precision.

It is also important to note that for these non-identified parameters, the posterior standard deviation does not decline to zero as the sample size increases. In fact, in sufficiently large samples, the marginal posteriors will behave as the conditional priors (evaluated at the MLE). For the causal effect BMI coefficient, for example, we calculate a posterior standard deviation of .08. This suggests that, under iid sampling, *one million* posterior simulations would be needed to achieve this level of precision. Our method, by exploiting the known conditional density in a Rao-Blackwell sense, even offers a significant improvement over this gold standard.

A lesson here, however, is that regardless of the method employed, accurate estimates of the posterior moments will require a large number of simulations, significantly larger than the numbers that we might see in treatment-response modeling of this sort. The partial identification of the model conspires to slow down mixing, and marginal posteriors for the non-identified parameters do not collapse, again leading to increased numerical uncertainty in the estimation of posterior moments. Our semi-analytic method significantly lessens these computational demands and mitigates the potential for inaccuracy, although it is still prudent to generate a large number of posterior simulations of ψ .

¹⁵This was roughly consistent with our own experience on our own PCs. Admittedly, this performance will vary widely across machines and over time as computing power increases, but we regard it as useful benchmark simply for illustration purposes. The calculations are also biased in our favor since our semi-analytic method requires evaluation of the conditional prior over a grid for each post-convergence simulation. The extra time required to do this, however, is mostly negligible and does not in any way alter the conclusions drawn from the table.

[Table 1 about here.]

In Table 2 we repeat the analysis of Table 1 after strengthening the prior by choosing $\alpha_{1i} \overset{iid}{\sim} N(0, .01^2)$, $i = 1, 2$. This, to us, seems to be a reasonable prior to adopt: we center the direct effects of the instruments over zero and regard it as unlikely that a one point in parental BMI will directly increase or decrease child earnings by more than 2 percent, given the set of controls that we employ (which include the BMI of the respondent). As we see from the table, the increased prior information improves the performances of both the standard method and our semi-analytic approach. The few cases for which the required number of iterations are observed to increase in Table 2 (relative to Table 1) under this tighter prior are those for which the posterior means have become closer to zero, thus requiring a higher precision standard for the entries in this second table.¹⁶ This improvement in performance is, of course, sensible, since the tighter prior leads us closer to a point identified model. However, even in this case with a reasonably strong prior, simply averaging the Gibbs simulations still proves to be potentially problematic. For the causal effect coefficient, coefficients associated with the parental BMI instruments and the error correlation, we would still require many millions of simulations in order to achieve the set level of accuracy. On the other hand, our semi-analytic approach can accomplish the same goal within a few hundred thousand iterations, and within one hour of computation time.

[Table 2 about here.]

4.1. Posterior Results and Prior Sensitivity. As we reanalyze the data of Kline and Tobias (2008) in further detail, we proceed as follows. We first obtain results, reported in Table 3, under the conventional assumption employed by Kline and Tobias (2008), that the instruments are completely excludable. We will then relax this assumption using our methods and thereby investigate the sensitivity of results to the imposition of this dogmatic prior belief.

Our results under the dogmatic prior (based on 45,000 post-convergence simulations) mimic those reported in Kline and Tobias (2008), essentially replicating their results.¹⁷ We and they find a clearly negative impact of BMI on earnings (in our analysis, a one point increase in BMI

¹⁶Said differently, had the posterior means in Table 2 remained constant, the required number of iterations would have decreased for all model parameters.

¹⁷Posterior means and standard deviations are quite close to those reported in Table II of Kline and Tobias (2008), although not identical. These authors, however, used a skew- t model for the wage equation, and treated the BMI-earnings relationship nonparametrically. These specification differences likely account for the slight discrepancy in posterior results.

lowers expected wages by approximately 1.8 percent on average, with $\Pr(\gamma < 0|\mathbf{y}, \mathbf{s}) \approx .99$, and endogeneity is important in the sense that $\Pr(\rho_{uv} > 0|\mathbf{y}, \mathbf{s}) \approx .98$. Kine and Tobias (2008) also separately examined the BMI-log wage relationships by BMI category (normal weight, overweight and obese) and obtained, for example, an average BMI penalty of 1.6 percent over the normal weight range. This, again, is broadly consistent with our results.

[Table 3 about here.]

4.2. Truncated Normal Priors for the IVs. We now seek to determine the sensitivity of posterior results in Table 3 to prior information concerning the excludability of the instruments. To this end, following (3), we begin by considering a variety of priors that formally impose the ordering of direct and indirect effects and some that truncate the support of the error correlation ρ_{uv} . The variants of (3) and (6) we consider are the following four:

- **P1** : $\alpha_{1i} | \gamma, \beta_{1i} \sim TN_{(-|\gamma\beta_{1i}|, |\gamma\beta_{1i}|)}(0, .008^2), i = 1, 2$, and $\rho_{uv} \sim TN_{I(-1 < \rho_{uv} < 1)}(0, .2^2)$
- **P2** : $\alpha_{1i} | \gamma, \beta_{1i} \sim TN_{(-|\gamma\beta_{1i}|, |\gamma\beta_{1i}|)}(0, 1), i = 1, 2$, and $\rho_{uv} \sim TN_{I(-1 < \rho_{uv} < 1)}(0, .2^2)$;
- **P3** : $\alpha_{1i} | \gamma, \beta_{1i} \sim TN_{(-|\gamma\beta_{1i}|, |\gamma\beta_{1i}|)}(0, 1), i = 1, 2$, and $\rho_{uv} \sim TN_{I(-1 < \rho_{uv} < 1)}(0, .5^2)$;
- **P4** : $\alpha_{1i} | \gamma, \beta_{1i} \sim TN_{(-|\gamma\beta_{1i}|, |\gamma\beta_{1i}|)}(0, 1), i = 1, 2$, and $\rho_{uv} \sim TN_{I(-.5 < \rho_{uv} < .5)}N(0, .5^2)$.

[Figure 3 about here.]

Marginal posteriors for the causal effect parameter, γ under these four different priors are reported in Figure 3. Prior **P1**, in the absence of the ordering restriction, puts a small prior standard deviation of .008 on the direct IV effects, suggesting it would be unlikely that a one point increase in parental BMI will directly increase/decrease child earnings by more than 1.6 percent. In addition, this prior also asserts that the error correlation is likely to be modest in value, placing approximately 95 percent of its prior mass over $(-.4, .4)$. As we move from **P2** to **P1**, we visualize the change in the causal effect posterior as a result of tightening this (untruncated) variance parameter from 1 to .008². We find that the posterior under **P2** is slightly more spread out than that under **P1**, but not dramatically so, as most of the information contained in the posterior is coming through the ordering restriction itself. **P3** might be interpreted as saying very little *a priori*, apart from simply ordering the direct and indirect effects. With the adoption of this relatively weak prior information, the posterior spreads out over a much wider range and places much more mass over large negative values, clearly illustrating the impact of the prior on posterior results. Analysis of results under **P4** then reveals how much is learned relative to **P3** when restricting the error correlation to lie in the interval $(-.5, .5)$. As the figure shows, this truncation of the error correlation results in a substantial tightening of the causal effect

posterior relative to that obtained under **P3**, and the posterior becomes flatter over the interval $(-.06, -.02)$. This suggests that priors truncating the error term can lead to much sharper posteriors; in our view, ruling out very large values of the error correlation is sensible in this application and may be sensible in many applications, given the large set of controls that are commonly employed.

All of the marginal posteriors in Figure 3 place all their mass over negative values and no mass over zero. As mentioned in section 2.1 and revealed in the conditional prior following equation (10), this is a consequence of the ordering restriction in the prior. We thus consider results under a variety of other priors to see if the finding of a negative impact of BMI on earnings is maintained.

First, Table 4 reports parameter posterior means and related features when adopting priors similar to those in Conley et al (2012). Specifically, we obtain results when assuming $\alpha_{1i} \stackrel{iid}{\sim} N(0, .008^2)$, and keeping the remaining priors like those used in the experiments at the beginning of this section. Here, we find several interesting results. First, point estimates of parameters in the BMI equation in Table 4 have not significantly changed relative to those reported in Table 3, which is to be expected given that these parameters are identifiable and thus minimally influenced by the change in prior. Second, despite the fact that our prior for the causal effect parameter is symmetric and centered over zero, the posterior shifts to place most of its mass over negative values, with $\Pr(\gamma < 0 | \mathbf{y}, \mathbf{s}) \approx .85$. Thus, we would continue to argue that BMI has a negative impact on female earnings, even when allowing for this uncertainty over instrument validity and not taking a stand *a priori* on the sign of the causal impact. In terms of point estimates, the posterior mean for the causal effect under this prior (-.0169) is quite similar to that obtained when the IVs are assumed to be perfectly excludable (-.018) although the posterior standard deviation for this parameter has more than tripled relative to that reported in Table 3. This loss in precision is to be expected as a natural consequence of the reduced prior information and the ensuing partial identification of the model.

[Table 4 about here.]

Tables 5 and 6 provide results under priors like those in equations (4) and (5), respectively. In Table 5, we report results under the priors:

$$\alpha_{1i} | \gamma^*, \beta_{1i} \stackrel{iid}{\sim} TN_{(-|\gamma^* \beta_{1i}|, |\gamma^* \beta_{1i}|)}(0, .008^2), \quad i = 1, 2$$

where we choose $\gamma^* = -.025$. As discussed in section 2, one possible way of eliciting the value of γ^* is to think about something akin to a maximal causal effect, so that the resulting prior can

be interpreted as approximately restricting the range of the direct IV impacts α_{1i} to be smaller than the maximum value of the indirect effect. The identifiable parameter β_1 is introduced into the prior to think about things in this way, and thereby to aid in eliciting hyperparameters of the prior. As shown in Table 5, results here are again very similar to those obtained previously, with a causal effect posterior mean equal to $-.0173$. In addition, this prior leads to a tighter posterior distribution for the causal effect parameter (relative to Table 4) as $\Pr(\gamma < 0 | \mathbf{y}, \mathbf{s}) \approx .93$.

[Table 5 about here.]

Table 6 considers the hierarchical prior specification

$$\alpha_{1i} | \gamma^*, \beta_{1i} \stackrel{iid}{\sim} TN_{(-|\gamma^* \beta_{1i}|, |\gamma^* \beta_{1i}|)}(0, .008^2), \quad \text{with} \quad \gamma^* \sim TN_{(0, .05)}(.025, .01^2).$$

Results under this prior are, again, very close to those reported in Table 5 and similar to those obtained under complete excludability.

[Table 6 about here.]

These results of this reexamination clearly indicate that priors will matter in terms of the causal effect posterior distributions. This is necessarily the case, given the partial identification of the model when non-dogmatic priors are employed. What we argue here is that it seems desirable to provide a menu of posterior results in these types of analyses which correspond to the adoption of different types of prior information. The viewpoint that the instruments are perfectly excludable simply represents a single item on this menu - an item that may be unpalatable to some. When assigning reasonable limits on the likely extent of instrument imperfection, or formally / approximately ordering the direct and indirect effects, our results continue to provide evidence of a negative impact of BMI on female earnings. Finally, our semi-analytic approach enables us to calculate these effects from the partially identified model reasonably quickly and reliably.

5. CONCLUSION

In the context of the triangular linear SEM, we have considered the issue of instrument imperfection. Instead of imposing that the instruments are excludable, we, like others have recently done in the literature, introduced a prior that allows the instruments to directly influence the outcome, but impose that this effect must be small. Our notion of “small” here is primarily a relative one, as our prior imposes the belief that the direct effect of the instrument must be

smaller than the indirect effect of the instrument through its intermediate relationship with the endogenous variable.

In addition to introducing and motivating this prior, we also documented issues surrounding posterior computation. Specifically, we showed that the non-identified simulations mix very poorly, and as a consequence, posterior statistics based on direct use of those simulations can prove unreliable. As an alternative, we derived a semi-analytic approach that characterized the conditional priors analytically and used these results in conjunction with the well-mixing identified posterior simulations to perform the calculations.

Finally, we illustrated our methods in studying the effect of BMI on earnings, using the data set employed by Kline and Tobias (2008). We approximately replicated their results when excludability is assumed, and obtained a set of new results when that viewpoint is relaxed. The different priors employed all seemed to point toward the same conclusion that BMI had a negative impact on female earnings.

We hope the general idea of providing a range of posterior results under different types of prior beliefs continues to become more popular in these types of studies. In addition, we plan to continue working on different types of models, such as nonlinear ones, where the semi-analytic scheme can be applied, as well as thinking about alternate numerical strategies, as briefly mentioned in the paper, when the analytic derivations are difficult to come by or simply not available.

6. APPENDIX

Section 3 presents a method for calculating (up to proportionality) the conditional prior $p(\gamma|\boldsymbol{\psi})$, and describes the results of this calculation in detail. This approach can be applied to all non-identified parameters, and marginal posterior statistics for each of these can be obtained from this more efficient procedure. Doing so requires repeating the change of variables for each new parameter under consideration, deriving the inverse transformations and the associated Jacobians of the transformations. We list key results for those exercises below. We, again, assume that the prior in (3) has been adopted, although we have derived similar results for the priors in (4) and (5).

6.1. $p(\alpha_i|\boldsymbol{\psi})$:

With a slight abuse of notation, we consider here derivation of a conditional prior of the form $p(\alpha_i|\boldsymbol{\psi})$. Although $\boldsymbol{\alpha}_2$ is a vector, the result given here applies to any specific element of $\boldsymbol{\alpha}_2$, or directly to either the intercept α_0 or the instrumental variable coefficient α_1 . Thus, when we index the parameters by i , we mean $i \in \{0, 1\}$ or that i simply indexes the individual elements of $\boldsymbol{\alpha}_2$.

Here, we consider the transformation from $\boldsymbol{\theta} = [\alpha_i \sigma_v^2 \beta_0 \beta_1 \boldsymbol{\beta}_2 \boldsymbol{\alpha}_{-i} \gamma \sigma_u^2]$ to $\boldsymbol{\psi}_{\alpha_i} = [\alpha_i \sigma_v^2 \beta_0 \beta_1 \boldsymbol{\beta}_2 \boldsymbol{a}_{-i} b c^2]$, with \boldsymbol{x}_{-i} denoting all elements of \boldsymbol{x} other than the i^{th} . One can show, following the arguments presented in section 3, that the non-trivial inverse transformations in this case are of the forms:

$$\begin{aligned} \boldsymbol{\alpha}_{-i} &= \mathbf{g}_{\boldsymbol{\alpha}_{-i}}^{-1}(\boldsymbol{\psi}_{\alpha_i}) = \boldsymbol{a}_{-i} + \boldsymbol{\beta}_{-i} \beta_0^{-1} (\alpha_i - a_i) \\ \gamma &= g_{\gamma}^{-1}(\boldsymbol{\psi}_{\alpha_i}) = b - \beta_0^{-1} (\alpha_i - a_i) \\ \sigma_u^2 &= g_{\sigma_u^2}^{-1}(\boldsymbol{\psi}_{\alpha_i}) = c^2 + \sigma_v^2 \beta_0^{-2} (\alpha_i - a_i)^2 \\ \rho_{uv} &= g_{\rho}^{-1}(\boldsymbol{\psi}_{\alpha_i}) = \frac{\sigma_v \beta_0^{-1} (\alpha_i - a_i)}{\sqrt{c^2 + (\alpha_i - a_i)^2 \sigma_v^2 \beta_0^{-2}}} \end{aligned}$$

This leads to the following Jacobian of the transformation:

$$\frac{\sigma_u |\beta_i^{-1}|}{c^2 + (\alpha_i - a_i)^2 \sigma_v^2 \beta_i^{-2}}$$

and thus

$$p(\alpha_i|\boldsymbol{\psi}) \propto p(\boldsymbol{\psi}_{\alpha_i}) = \frac{\sigma_v |\beta_i^{-1}|}{c^2 + (\alpha_i - a_i)^2 \sigma_v^2 \beta_i^{-2}} p[\mathbf{g}_{\boldsymbol{\alpha}_{-i}}^{-1}(\boldsymbol{\psi}_{\alpha_i})].$$

The methods discussed in section (3) can then be used to calculate marginal posterior statistics for α_i .

6.2. $p(\rho_{uv}|\boldsymbol{\psi})$:

To derive the kernel of this conditional prior, we consider the transformation from $\boldsymbol{\theta} = [\rho_{uv} \sigma_v^2 \beta_0 \beta_1 \boldsymbol{\beta}_2 \alpha_0 \alpha_1 \boldsymbol{\alpha}_2 \gamma \sigma_u^2]$ to $\boldsymbol{\psi}_\rho = [\rho_{uv} \sigma_v^2 \beta_0 \beta_1 \boldsymbol{\beta}_2 a_0 a_1 \boldsymbol{a}_2 b c^2]$: In this case, the non-trivial inverse mappings are of the forms:

$$\begin{aligned}\alpha_0 &= g_{\alpha_0}^{-1}(\boldsymbol{\psi}_\rho) = a_0 + \beta_0 \sigma_v^{-1} \sqrt{\frac{c^2}{1 - \rho_{uv}^2}} \\ \alpha_1 &= g_{\alpha_1}^{-1}(\boldsymbol{\psi}_\rho) = a_1 + \beta_1 \sigma_v^{-1} \sqrt{\frac{c^2}{1 - \rho_{uv}^2}} \\ \boldsymbol{\alpha}_2 &= \mathbf{g}_{\boldsymbol{\alpha}_2}^{-1}(\boldsymbol{\psi}_\rho) = \boldsymbol{a}_2 + \boldsymbol{\beta}_2 \sigma_v^{-1} \sqrt{\frac{c^2}{1 - \rho_{uv}^2}} \\ \gamma &= g_\gamma^{-1}(\boldsymbol{\psi}_\rho) = b - \sigma_v^{-1} \rho_{uv} \sqrt{\frac{c^2}{1 - \rho_{uv}^2}} \\ \sigma_u^2 &= g_{\sigma_u^2}^{-1}(\boldsymbol{\psi}_\rho) = \frac{c^2}{1 - \rho_{uv}^2}\end{aligned}$$

This leads to the following Jacobian of the transformation:

$$(1 - \rho_{uv}^2)^{-1}$$

and thus

$$p(\rho_{uv}|\boldsymbol{\psi}) \propto p(\boldsymbol{\psi}_\rho) = (1 - \rho_{uv}^2)^{-1} p[\mathbf{g}_\rho^{-1}(\boldsymbol{\psi}_\rho)].$$

6.3. $p(\sigma_u^2|\boldsymbol{\psi})$:

Derivation of this conditional prior is slightly more complicated than the others considered thus far. To establish this conditional prior up to proportionality, we choose to first reparameterize ρ_{uv} as:

$$\rho_{uv} = \text{sign}(\rho_{uv}) |\rho_{uv}| \equiv \zeta \phi,$$

so that $\zeta \in \{-1, 1\}$ and $\phi \geq 0$. If we employ the following independent priors,

$$\mathbb{P}(\zeta = -1) = \mathbb{P}(\zeta = 1) = 0.5, \quad \phi \sim TN_{(0,1)}(0, \underline{V}_\rho).$$

then the induced prior on ρ_{uv} is

$$\rho_{uv} \sim TN_{(-1,1)}(0, \underline{V}_\rho),$$

similar to the prior described in (4).¹⁸

Let $\tilde{\boldsymbol{\theta}} = (\sigma_u^2, \sigma_v^2, \beta_0, \beta_1, \boldsymbol{\beta}'_2, \alpha_0, \alpha_1, \boldsymbol{\alpha}'_2 \gamma, \phi)'$ and write the prior in $(\zeta, \tilde{\boldsymbol{\theta}})$ parameterization as

$$p(\tilde{\boldsymbol{\theta}} | \zeta) p(\zeta).$$

Since

$$p(\sigma_u^2 | \boldsymbol{\psi}) \propto p(\sigma_u^2, \boldsymbol{\psi}) = 0.5p(\sigma_u^2, \boldsymbol{\psi} | \zeta = -1) + 0.5p(\sigma_u^2, \boldsymbol{\psi} | \zeta = 1),$$

it suffices to derive an expression to evaluate $p(\sigma_u^2, \boldsymbol{\psi} | \zeta)$. As will be shown below, conditioning on ζ , the sign of ρ_{uv} , facilitates the change of variables exercise and motivates our reparameterization.

We consider the transformation $\mathbf{g}_{\sigma_u^2} : \tilde{\boldsymbol{\theta}} \mapsto \boldsymbol{\psi}_{\sigma_u^2} = [\sigma_u^2 \ \sigma_v^2 \ \beta_0 \ \beta_1 \ \boldsymbol{\beta}_2 \ a_0 \ a_1 \ \mathbf{a}_2 \ b \ c^2]$ (throughout ζ is fixed). The non-trivial inverse mappings are of the forms:

$$\begin{aligned} \alpha_0 &= g_{\alpha_0}^{-1}(\boldsymbol{\psi}_{\sigma_u^2}) = a_0 + \zeta \sqrt{\frac{\sigma_u^2 - c^2}{\sigma_v^2}} \beta_0, \\ \alpha_1 &= g_{\alpha_1}^{-1}(\boldsymbol{\psi}_{\sigma_u^2}) = a_1 + \zeta \sqrt{\frac{\sigma_u^2 - c^2}{\sigma_v^2}} \beta_1, \\ \boldsymbol{\alpha}_2 &= g_{\boldsymbol{\alpha}_2}^{-1}(\boldsymbol{\psi}_{\sigma_u^2}) = \mathbf{a}_2 + \zeta \sqrt{\frac{\sigma_u^2 - c^2}{\sigma_v^2}} \boldsymbol{\beta}_2, \\ \gamma &= g_{\gamma}^{-1}(\boldsymbol{\psi}_{\sigma_u^2}) = b - \zeta \sqrt{\frac{\sigma_u^2 - c^2}{\sigma_v^2}}, \\ \phi &= g_{\phi}^{-1}(\boldsymbol{\psi}_{\sigma_u^2}) = \sqrt{1 - \frac{c^2}{\sigma_u^2}}. \end{aligned}$$

The Jacobian of this transformation can be shown to equal:

$$|\det \mathbf{J}(\boldsymbol{\psi}_{\sigma_u^2})| = \left| \frac{\partial \phi}{\partial c^2} \right| = \frac{1}{2\sigma_u \sqrt{\sigma_u^2 - c^2}}.$$

The conditional prior density can therefore be calculated up to proportionality:

$$p(\sigma_u^2 | \boldsymbol{\psi}) \propto p(\sigma_u^2, \boldsymbol{\psi}) = 0.5[p(\sigma_u^2, \boldsymbol{\psi} | \zeta = -1) + p(\sigma_u^2, \boldsymbol{\psi} | \zeta = 1)],$$

where

$$p(\sigma_u^2, \boldsymbol{\psi} | \zeta) = |\mathbf{J}(\boldsymbol{\psi}_{\sigma_u^2})| p(\mathbf{g}_{\sigma_u^2}^{-1}(\boldsymbol{\psi}_{\sigma_u^2}) | \zeta).$$

¹⁸It is of course possible to construct in a similar way an induced prior of the form $TN_{(\rho_l, \rho_u)}(\underline{\mu}_\rho, \underline{V}_\rho)$. We consider the above prior for illustration purposes.

REFERENCES

- ANGRIST, J., AND A. KRUEGER (1991): “Does Compulsory School Attendance Affect Schooling and Earnings?,” *Quarterly Journal of Economics*, pp. 979–1014.
- ASHLEY, R. (2009): “Assessing the Credibility of Instrumental Variables Inference with Imperfect Instruments via Sensitivity Analysis,” *Journal of Applied Econometrics*, 24, 325–337.
- BUCKLES, K., AND D. HUNGERMAN (2010): “Season of Birth and Later Outcomes: Old Questions, New Answers,” .
- CAWLEY, J. (2004): “The Impact of Obesity on Wages,” *Journal of Human Resources*, 39, 451–474.
- CHIB, S., AND B. HAMILTON (2000): “Bayesian Analysis of Cross-Section and Clustered Data Selection Models,” *Journal of Econometrics*, pp. 25–50.
- (2004): “Semiparametric Bayes Analysis of Longitudinal Data Treatment Models,” *Journal of Econometrics*, pp. 67–89.
- CONLEY, T.G., . C. B. H., AND P. ROSSI (2012): “Plausibly Exogenous,” *Review of Economics and Statistics*, pp. 260–272.
- DEARDEN L., F. F., AND C. MEGHIR (2002): “The Effect of School Quality on Educational Attainment and Wages,” *Review of Economics and Statistics*, 84(1), 1–20.
- DEB, P., M. K. MUNKIN, AND P. K. TRIVEDI (2006): “Bayesian Analysis of the Two-Part Model with Endogeneity: Application to Health Care Expenditure,” *Journal of Applied Econometrics*, 21, 1081–1099.
- GEWEKE, J., G. GOWRISANKARAN, AND R. J. TOWN (2003): “Bayesian Inference for Hospital Quality in a Selection Model,” *Econometrica*, 71(4), 1215–1283.
- GUNDERSEN, C., AND B. KREIDER (2009): “Bounding the Effects of Food Insecurity on Children’s Health Outcomes,” *Journal of Health Economics*, 28, 971–983.
- KLINE, B., AND J. TOBIAS (2008): “The Wages of BMI: Bayesian Analysis of a Skewed Treatment-Response Model with Nonparametric Endogeneity,” *Journal of Applied Econometrics*, 23(6), 767–793.
- KOOP, G., AND D. POIRIER (1997): “Learning About the Across-Regime Correlation in Switching Regression Models,” *Journal of Econometrics*, pp. 217–227.
- KRAAY, A. (2012): “Instrumental Variables Regressions with Uncertain Exclusion Restrictions: A Bayesian Approach,” *Journal of Applied Econometrics*, 27, 108–128.
- LI, K. (1998): “Bayesian Inference in a Simultaneous Equation Model with Limited Dependent Variables,” *Journal of Econometrics*, 85(2), 387–400.
- LI, K., AND D. POIRIER (2003a): “The Roles of Birth Inputs and Outputs in Predicting Health, Behavior, and Test Scores in Early Childhood,” *Statistics in Medicine*.
- (2003b): “An Econometric Model of Birth Weight for Native Americans,” *Journal of Econometrics*.
- MANSKI, C., AND J. PEPPER (2000): “Monotone Instrumental Variables: With an Application to the Returns to Schooling,” *Econometrica*, 68(4), 997–1010.

- MOON, H., AND F. SCHORFHEIDE (2011): “Bayesian and Frequentist Inference in Partially Identified Models,” *Econometrica* (forthcoming).
- MUNKIN, M. K., AND P. K. TRIVEDI (2003): “Bayesian Analysis of a Self-Selection Model with Multiple Outcomes Using Simulation-Based Estimation: An Application to the Demand for Health Care,” *Journal of Econometrics*, 114(2), 197–220.
- (2008): “Bayesian Analysis of the Ordered Probit Model with Endogenous Selection,” *Journal of Econometrics*, 143(2), 334–348.
- NEVO, A., AND A. ROSEN (2010): “Identification with Imperfect Instruments,” *Review of Economics and Statistics* (forthcoming).
- POIRIER, D. (1998): “Revising Beliefs in Nonidentified Models,” *Econometric Theory*, pp. 482–509.

FIGURE 1. Lagged Autocorrelations for a sample of Non-Identified Parameters.

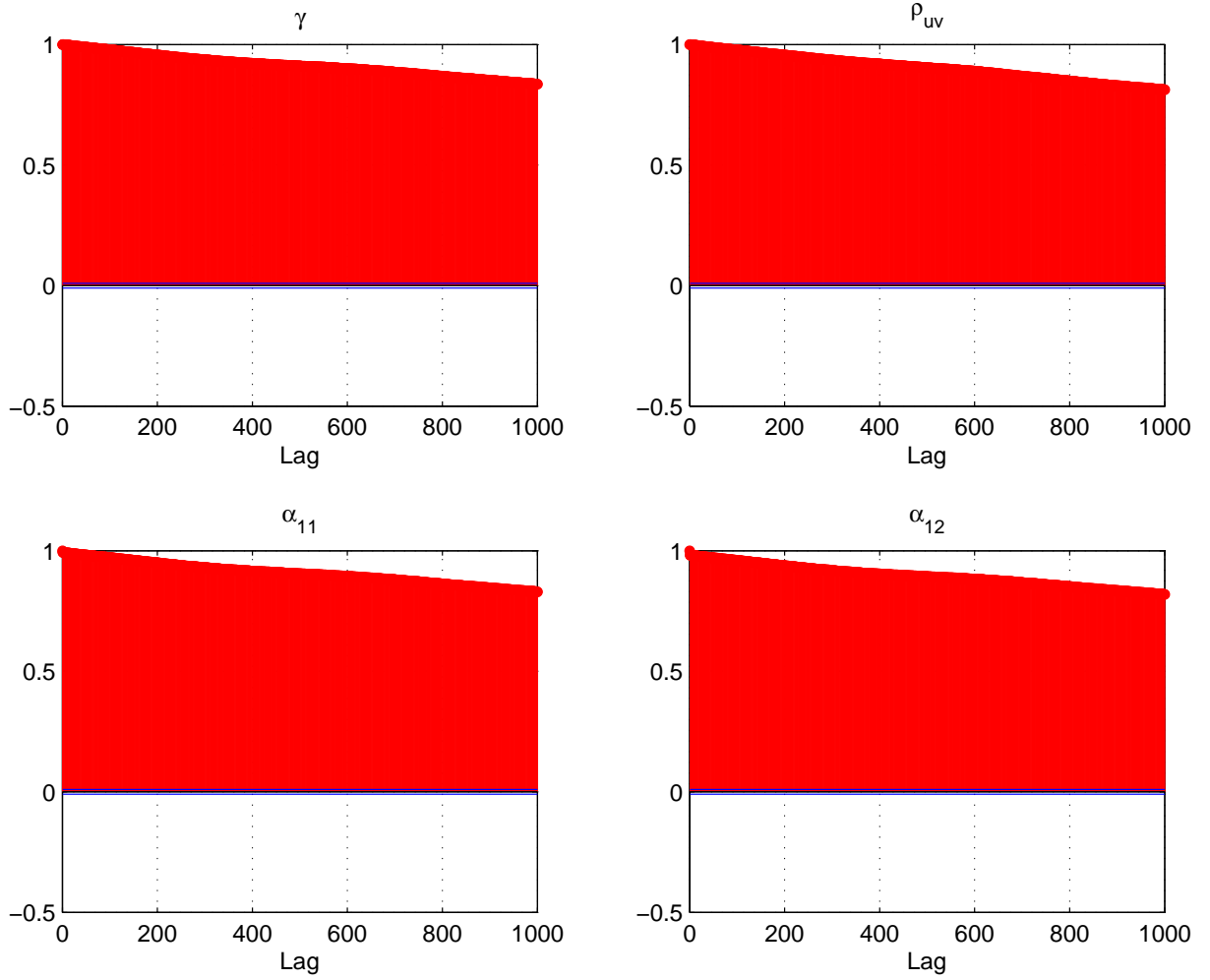


FIGURE 2. Lagged Autocorrelations for a sample of Identified Parameters.

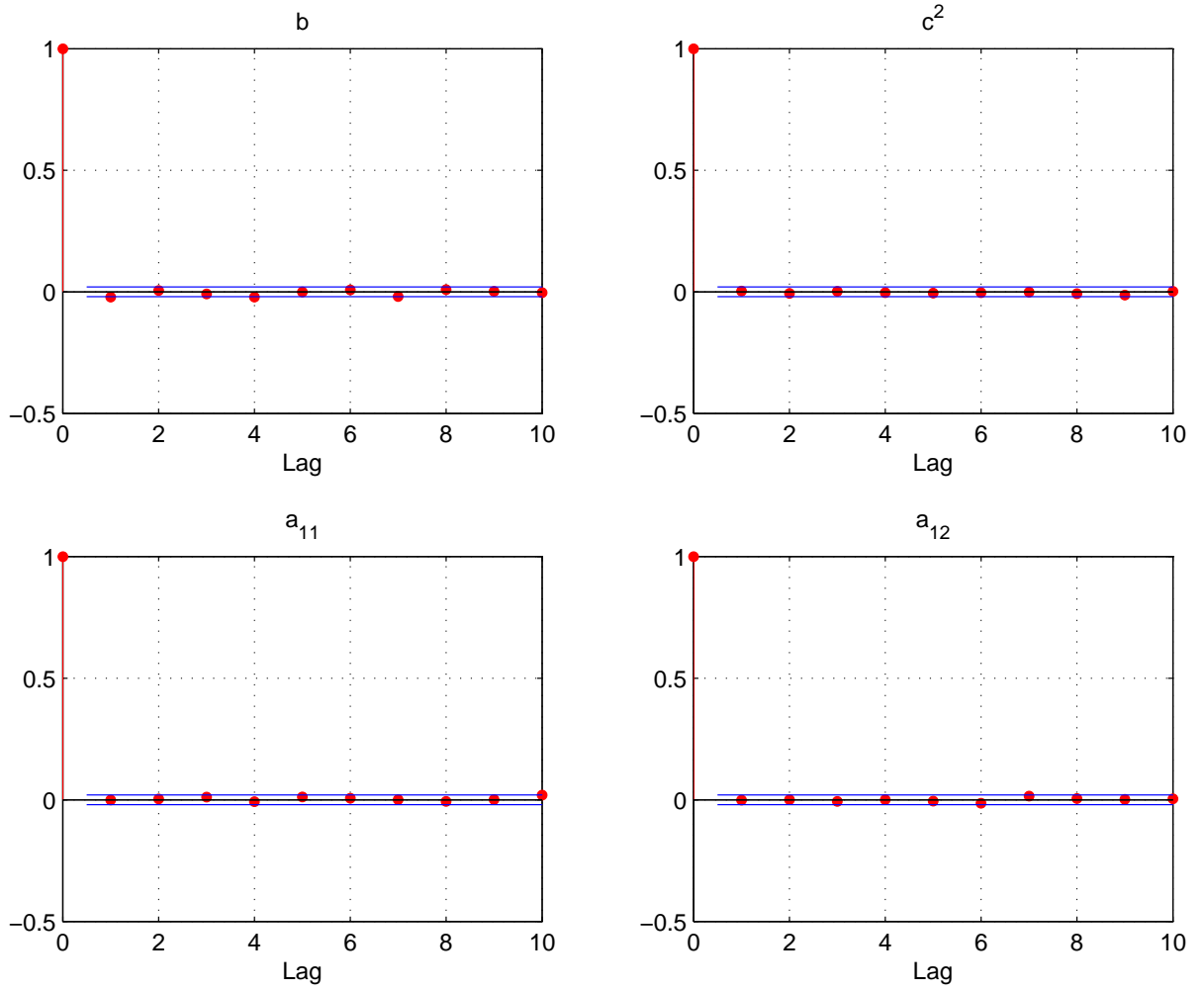


FIGURE 3. The marginal posterior densities $p(\gamma | \mathbf{y},)$ under various prior assumptions: P1 (left top), P2 (right top), P3 (left bottom) and P4 (right bottom).

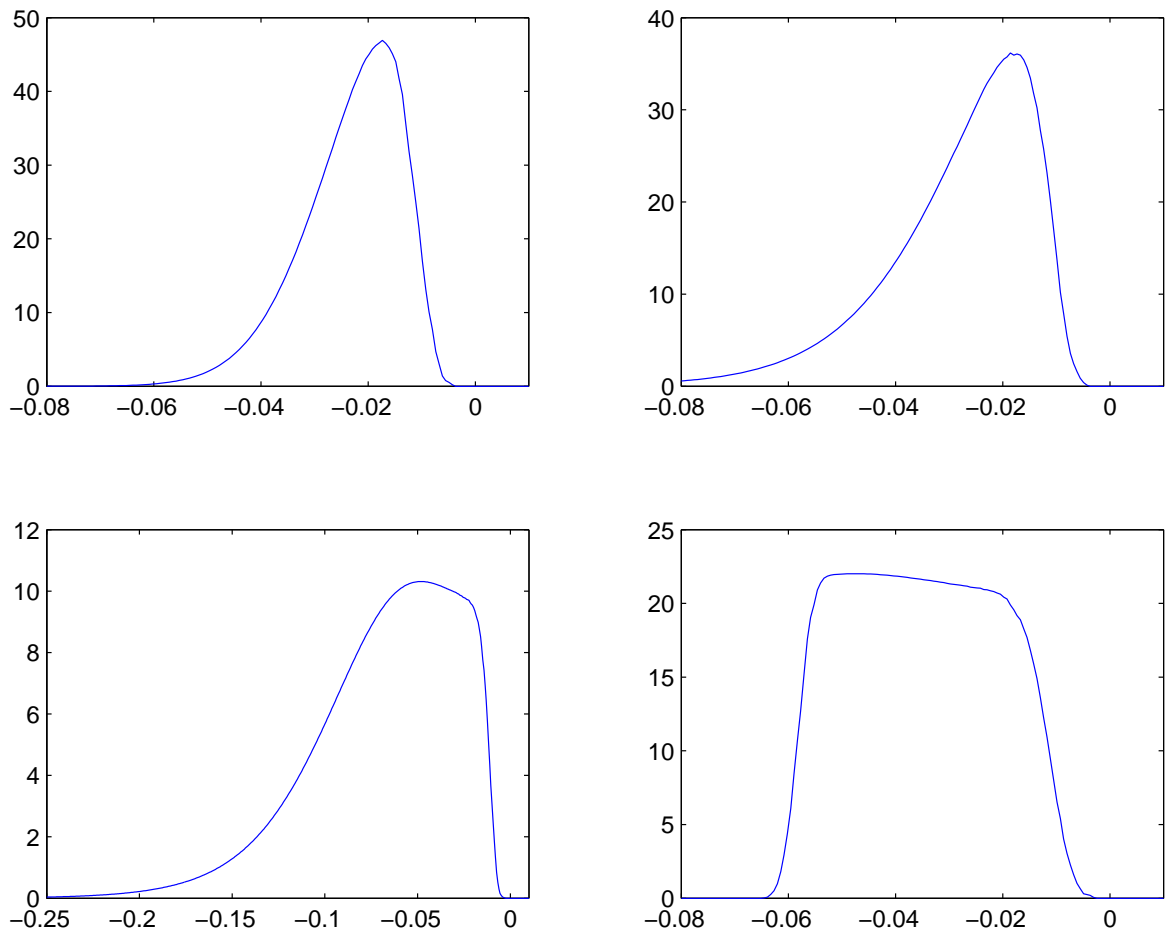


TABLE 1. Number of draws (in thousands) required to have the NSE less than one percent of the posterior mean: $\alpha_{1i} \sim N(0, .05^2)$.

Wage Equation					
Parameter	$E(\cdot Data)$	Number of Iterations		Computation Time	
		Gibbs	Semi-Analytic	Gibbs	Semi-Analytic
BMI	.008	3,377,208	146	390 days	24 minutes
Constant	1.756	5,685	1.50	15.8 hours	.25 minutes
MomBMI	-.0092	363,582	8	42.1 days	1.2 minutes
DadBMI	-.0073	317,180	24	36.7 days	4.0 minutes
FamilyIncome	.0007	218	63	0.6 hours	10 minutes
HighSchool	.062	5,659	14	15.7 hours	2.4 minutes
Alevel	.266	899	1.92	2.5 hours	.32 minutes
Degree	.355	903	.58	2.5 hours	.09 minutes
Union	.031	6,852	38	19.0 hours	6.3 minutes
Married	-.018	55,708	106	6.4 days	17 minutes
Other Parameters					
Parameter	$E(\cdot Data)$	Gibbs	Semi-Analytic	Gibbs	Semi-Analytic
ρ_{uv}	-.097	965,848	5	112 days	.90 minutes
σ_u^2	.225	12,293	.03	1.4 days	.005 minutes

TABLE 2. Number of draws (in thousands) required to have the NSE less than one percent of the posterior mean: $\alpha_{1i} \sim N(0, .01^2)$.

Wage Equation					
Parameter	$E(\cdot Data)$	Number of Iterations		Computation Time	
		Gibbs	Semi-Analytic	Gibbs	Semi-Analytic
BMI	-.016	14,516	35	40.3 hours	5.8 minutes
Constant	1.965	27	.52	4.5 minutes	.08 minutes
MomBMI	-.0006	1,253,226	282	145 days	47 minutes
DadBMI	-.0008	338,364	391	39.1 days	65 minutes
FamilyIncome	.0007	4	3	0.6 minutes	.58 minutes
HighSchool	.070	72	4	12 minutes	.65 minutes
Alevel	.278	13	.39	2.1 minutes	.06 minutes
Degree	.339	18	.38	2.9 minutes	.06 minutes
Union	.035	104	14	17 minutes	2.3 minutes
Married	-.012	2,464	99	6.8 hours	17 minutes
Other Parameters					
Parameter	$E(\cdot Data)$	Gibbs	Semi-Analytic	Gibbs	Semi-Analytic
ρ_{uv}	.092	47,041	130	5.4 days	22 minutes
σ_u^2	.133	52	0.06	8.6 minutes	.01 minutes

TABLE 3. Parameter posterior means, standard deviations and probabilities begin positive. Exclusion restrictions imposed.

Wage Equation			
Variable	$\mathbb{E}(\cdot \mathbf{y})$	$\sqrt{\text{Var}(\cdot \mathbf{y})}$	$\mathbb{P}(\cdot > 0 \mathbf{y})$
Constant	1.989	.1477	1.00
JobTenure	.0212	.0075	.9978
JobTenure ²	-.0009	.0006	.0487
Experience	.0243	.0150	.9522
Experience ²	-.001	.0008	.1103
FamilyIncome	.0007	.0002	1.00
HighSchool	.0707	.0226	.9992
ALevel	.2798	.0346	1.00
Degree	.3378	.0309	1.00
Union	.0357	.0194	.9675
Married	-.0111	.0179	.2642
MomDegree	.0629	.0500	.8983
MomManProf	-.0062	.0253	.4027
DadDegree	-.0074	.0278	.3908
DadManProf	.0566	.0222	.9948
BMI	-.0182	.0051	.0003
BMI Equation			
Variable	$\mathbb{E}(\cdot \mathbf{y})$	$\sqrt{\text{Var}(\cdot \mathbf{y})}$	$\mathbb{P}(\cdot > 0 \mathbf{y})$
Constant	8.616	.9260	1.00
FamilyIncome	.0007	.0016	.6658
HighSchool	.3188	.2373	.9006
ALevel	.5212	.3509	.9378
Degree	-.7041	.2714	.0060
Union	.1683	.2078	.8026
Married	.2758	.1953	.9295
MomBMI	.3648	.0260	1.00
DadBMI	.2709	.0315	1.00
Other Parameters			
Variable	$\mathbb{E}(\cdot \mathbf{y})$	$\sqrt{\text{Var}(\cdot \mathbf{y})}$	$\mathbb{P}(\cdot > 0 \mathbf{y})$
ρ_{uv}	.1234	.0602	.9780
σ_u^2	.1270	.0047	1.00
σ_v^2	16.10	.5413	1.00

TABLE 4. Parameter posterior means, standard deviations and probabilities of being positive, $\alpha_{1i} \stackrel{iid}{\sim} N(0, .008^2)$, $i = 1, 2$.

Wage Equation			
Variable	$\mathbb{E}(\cdot \mathbf{y})$	$\sqrt{\text{Var}(\cdot \mathbf{y})}$	$\mathbb{P}(\cdot > 0 \mathbf{y})$
Constant	1.972	.2136	1.00
MomBMI	-.0004	.0064	.4948
DadBMI	-.0006	.0051	.4613
JobTenure	.0212	.0075	.9974
JobTenure ²	-.0009	.0006	.0507
Experience	.0244	.0150	.9487
Experience ²	-.0010	.0008	.1097
FamilyIncome	.0007	.0002	1.00
HighSchool	.0701	.0233	.9990
ALevel	.2789	.0361	1.00
Degree	.3387	.0337	1.00
Union	.0353	.0199	.9624
Married	-.0114	0.0186	.2776
MomDegree	.0629	.0498	.8956
MomManProf	-.0062	.0254	.4030
DadDegree	-.0080	.0277	.3858
DadManProf	.0566	.0222	.9949
BMI	-.0169	.0180	.1521
BMI Equation			
Variable	$\mathbb{E}(\cdot \mathbf{y})$	$\sqrt{\text{Var}(\cdot \mathbf{y})}$	$\mathbb{P}(\cdot > 0 \mathbf{y})$
Constant	8.638	.9279	1.00
FamilyIncome	.0007	.0016	.6722
HighSchool	.3067	.2377	.9026
ALevel	.5192	.3521	.9292
Degree	-.7045	.2717	.0050
Union	.1691	.2043	.7956
Married	.2760	.1919	.9242
MomBMI	.3649	.0262	1.00
DadBMI	.2703	.0317	1.00
Other Parameters			
Variable	$\mathbb{E}(\cdot \mathbf{y})$	$\sqrt{\text{Var}(\cdot \mathbf{y})}$	$\mathbb{P}(\cdot > 0 \mathbf{y})$
ρ_{uv}	.1032	.1913	.7260
σ_u^2	.1309	.0096	1.00
σ_v^2	16.085	0.5425	1.00

TABLE 5. Parameter posterior means, standard deviations and probabilities of begin positive: $\alpha_{1i}|\beta_{1i}, \gamma^* \stackrel{iid}{\sim} TN_{(-|\gamma^*\beta_{1i}|, |\gamma^*\beta_{1i}|)}(0, .008^2), i = 1, 2$, with $\gamma^* = -.025$.

Wage Equation			
Variable	$\mathbb{E}(\cdot \mathbf{y})$	$\sqrt{\text{Var}(\cdot \mathbf{y})}$	$\mathbb{P}(\cdot > 0 \mathbf{y})$
Constant	1.978	0.1755	1.00
MomBMI	-0.0002	0.0043	0.4781
DadBMI	-0.0005	0.0034	0.4416
JobTenure	0.0212	0.0076	0.9975
JobTenure ²	-0.0009	0.0006	0.0509
Experience	0.0245	0.0150	0.9505
Experience ²	-0.0010	0.0008	0.1095
FamilyIncome	0.0007	0.0002	1.00
HighSchool	0.0703	0.0229	0.9990
Alevel	0.2792	0.0353	1.00
Degree	0.3382	0.0321	1.00
Union	0.0352	0.0195	0.9628
Married	-0.0115	0.0179	0.2603
MomDegree	0.0634	0.0499	0.8984
MomManProf	-0.0062	0.0254	0.4061
DadDegree	-0.0078	0.0278	0.3895
DadManProf	0.0569	0.0222	0.9947
BMI	-0.0173	0.0118	0.0689
BMI Equation			
Variable	$\mathbb{E}(\cdot \mathbf{y})$	$\sqrt{\text{Var}(\cdot \mathbf{y})}$	$\mathbb{P}(\cdot > 0 \mathbf{y})$
Constant	8.590	0.9123	1.00
FamilyIncome	0.0007	0.0015	0.6707
HighSchool	0.3057	0.2367	0.9025
Alevel	0.5234	0.3500	0.9341
Degree	-0.7022	0.2742	0.0047
Union	0.1682	0.2071	0.7920
Married	0.2755	0.1922	0.9232
MomBMI	0.3653	0.0261	1.00
DadBMI	0.2718	0.0312	1.00
Other Parameters			
Variable	$\mathbb{E}(\cdot \mathbf{y})$	$\sqrt{\text{Var}(\cdot \mathbf{y})}$	$\mathbb{P}(\cdot > 0 \mathbf{y})$
ρ_{uv}	0.1111	0.1913	0.7770
σ_u^2	0.1288	0.0085	1.00
σ_v^2	16.089	0.5423	1.00

TABLE 6. Parameter posterior means, standard deviations and probabilities of being positive: $\alpha_{1i}|\gamma^*, \beta_{1i} \stackrel{iid}{\sim} TN_{(-|\gamma^*\beta_{1i}|, |\gamma^*\beta_{1i}|)}(0, .008^2)$, with $\gamma^* \sim TN_{(.05)}(.025, .01^2)$.

Wage Equation			
Variable	$\mathbb{E}(\cdot \mathbf{y})$	$\sqrt{\text{Var}(\cdot \mathbf{y})}$	$\mathbb{P}(\cdot > 0 \mathbf{y})$
Constant	1.983	0.1758	1.00
MomBMI	-0.0001	0.0042	0.4921
DadBMI	-0.0004	0.0032	0.4484
JobTenure	0.0213	0.0075	0.9976
JobTenure ²	-0.0009	0.0006	0.0498
Experience	0.0242	0.0149	0.9471
Experience ²	-0.0009	0.0008	0.1130
FamilyIncome	0.0007	0.0002	1.00
HighSchool	0.0705	0.0229	0.9991
ALevel	0.2801	0.0349	1.00
Degree	0.3381	0.0323	1.00
Union	0.0355	0.0196	0.9643
Married	-0.0110	0.0180	0.2706
MomDegree	0.0634	0.0495	0.9002
MomManProf	-0.0065	0.0255	0.4005
DadDegree	-0.0078	0.0278	0.3902
DadManProf	0.0567	0.0221	0.9955
BMI	-0.0178	0.0120	0.0677
BMI Equation			
Variable	$\mathbb{E}(\cdot \mathbf{y})$	$\sqrt{\text{Var}(\cdot \mathbf{y})}$	$\mathbb{P}(\cdot > 0 \mathbf{y})$
Constant	8.599	0.9222	1.00
FamilyIncome	0.0007	0.0016	0.6702
HighSchool	0.3076	0.2384	0.9007
ALevel	0.5219	0.3491	0.9323
Degree	-0.7048	0.2730	0.0053
Union	0.1678	0.2064	0.7903
Married	0.2770	0.1929	0.9240
MomBMI	0.3653	0.0261	1.00
DadBMI	0.2716	0.0311	1.00
Other Parameters			
Variable	$\mathbb{E}(\cdot \mathbf{y})$	$\sqrt{\text{Var}(\cdot \mathbf{y})}$	$\mathbb{P}(\cdot > 0 \mathbf{y})$
ρ_{uv}	0.1163	0.1272	0.8159
σ_u^2	0.1289	0.0066	1.00
σ_v^2	16.088	0.5416	1.00
c	0.0229	0.0096	1.00