

Conditional inference about the nonidentified correlation coefficient in the Roy model*

SAMITA SAREEN
Department of Economics
Duke University

Abstract

In the current literature on the extended Roy model of selectivity, marginal inference about the nonidentified correlation coefficient between the disturbance terms of the returns in the two sectors of choice is possible. However, data does not update prior beliefs about the nonidentified correlation coefficient, *conditional* on the identified parameters. This paper shows that if Jeffreys' prior based on the sampling distribution under *full observability* is used, prior beliefs about the nonidentified correlation coefficient, *conditional* on the identified parameters, are updated.

Keywords: Bayesian; Full observability; Jeffreys' prior; Partial observability; Roy model

JEL Classification: C11

*E-mail: ssareen@econ.duke.edu

1 INTRODUCTION

Roy's model of self-selection (1951) has been applied extensively in economics to explain individual choice between several alternatives or sectors. Assuming two sectors or alternatives, typically, in the extended or utility based version of the Roy model, two aspects of an agent's choice are observed. First, an agent's choice of an alternative and variables affecting the self-selection process are observed. Second, conditional on the choice, this agent's earnings or returns in the alternative of choice and variables affecting earnings in the several available alternatives, are observed. This shall be referred to as *partial observability* in this paper. *Full observability* will refer to the scenario, where in addition, an agent's earnings in the sectors she did not choose, are observed.

Since an agent's earnings are observed *only* in the sector of choice under *partial observability*, the correlation coefficient between the disturbance terms in the earnings equations is not identified under *partial observability*. Several papers have pointed out why inference about the nonidentified correlation coefficient might be an interesting exercise (Heckman and Honore, 1990; Vijverberg, 1993; Koop and Poirier, 1997), and have suggested ways in which this might be accomplished.

The view held in the literature on Bayesian inference of the extended Roy model, when observed data corresponds to the *partial observability* case, is that marginal inference about the nonidentified correlation coefficient is possible. But data do not modify prior beliefs about the nonidentified correlation coefficient *conditional* on the parameters identified under *partial observability*.

Marginal inference is interesting if the aim is to ascertain the correlation between unobserved productivity in the two sectors *irrespective* of the factors that determine self-selection by an agent. For example, if productive personal characteristics, such as punctuality and dilligence, enhance an agent's productivity, they will do so irrespective of the sector chosen by this agent.

In several instances inference about the nonidentified correlation coefficient *conditional* on the parameters of the self-selection mechanism is of interest; the latter are identified under *partial observability*. For example, one may like to know whether unobserved factors affecting productivity in the two sectors are correlated conditional on an agent self-selecting himself to Sector 1. A value close to +1 will indicate that the set of occupation-specific skills that lead this agent to select sector 1, also make him more productive in Sector 2. A value close to -1 will indicate that the occupation-specific skills that lead this agent to select Sector 1, make him less productive in Sector 2.

The prior that has been the focus in the literature on Bayesian inference of the extended Roy model, is the uniform prior based on the positive definiteness of the covariance matrix of the data under *full observability*. A prior that has been ignored in this literature is Jeffrey's prior based on the *full observability* likelihood. This paper attempts to remedy the gap. I shall show that when the

partial observability likelihood function is combined with Jeffreys’ prior based on the of the *full observability* likelihood, prior beliefs about the nonidentified correlation coefficient conditional on parameters identified under *partial observability*, are updated. As a result, inference about the nonidentified correlation coefficient *conditional* on the parameters of the self-selection mechanism, is possible.

Section 2 describes the extended Roy model under *full observability* and *partial observability*. Section 3 obtains Jeffreys’ prior for the nonidentified correlation coefficient based on the *full observability* likelihood. The posterior for the correlation coefficient, conditional on the parameters of the self-selection mechanism, when the data corresponds to the *partial observability* case, is obtained as well. Section 4 concludes with a discussion.

2 The extended Roy model of selectivity

In the discussion that follows, random variables will be indicated by capital letters, and the realization of a random variable by a lower case letter; in addition, matrices will be indicated by bold faced letters.

$j = 1, \dots, n$ individuals face a choice between two “sectors”.¹ The potential earnings of each individual i in sectors 1 and 2 is given by Y_{1j}^* and Y_{2j}^* , respectively. \mathbf{X}_{1j} and \mathbf{X}_{2j} are observed covariates affecting potential earnings in sectors 1 and 2, respectively. These could be personal characteristics measuring skill-levels in each sector, as well as, sector specific characteristics. The unobserved factors affecting potential earnings in sector 1 and 2 are given by $\epsilon = [\epsilon_{1j}, \epsilon_{2j}]$. These unobserved factors are correlated across sectors with ρ_{12} indicating the correlation coefficient between these unobserved factors affecting potential earnings in the two sectors.

The self-selection mechanism is described by the indicator variable I_j^* . I_j^* indicates the difference in utility experienced by individual j when she works in sector 1 versus sector 2. This is affected by observed covariates \mathbf{W}_j and unobserved factors u_j . If $I_j^* \geq 0$, individual j will choose sector 1; if $I_j^* < 0$, individual j will choose sector 2.²

Omitting subscripts for individuals, the Roy model can be described by

$$I^* = \mathbf{W}\gamma + u, \quad Y_1^* = \mathbf{X}_1\beta_1 + \epsilon_1, \quad Y_2^* = \mathbf{X}_2\beta_2 + \epsilon_2, \quad (1)$$

where I^*, Y_1^*, Y_2^* are scalars. γ, β_1, β_2 are of dimension $k_w \times 1, k_1 \times 1$ and $k_2 \times 1$, respectively. $\mathbf{W}, \mathbf{X}_1, \mathbf{X}_2$ are of dimension $1 \times k_w, 1 \times k_1$, and $1 \times k_2$, respectively.

¹The term “sectors” is used broadly here. If occupational choice is the issue, then this could refer to the two sectors of employment. In the program evaluation literature the two “sectors” correspond to participating in a program or not.

²Roy (1951) assumed that an individual’s sectoral choice is based exclusively on sectoral earnings.

Following convention I shall assume $j = 1, \dots, n$ independently and identically distributed individuals, where for each individual j ,³

$$(I^*, Y_1^*, Y_2^*)' \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Omega}). \quad (2)$$

$N_3(\boldsymbol{\mu}, \boldsymbol{\Omega})$ indicates a trivariate Normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Omega}$. While $\boldsymbol{\mu}$ varies across the j individuals, $\boldsymbol{\Omega}$ does not. From Vijverberg (1993, p. 71),

$$\boldsymbol{\mu} = [\mathbf{W}\boldsymbol{\gamma}, \mathbf{X}_1\boldsymbol{\beta}_1, \mathbf{X}_2\boldsymbol{\beta}_2]' \quad \text{and} \quad \boldsymbol{\Omega} = \begin{bmatrix} 1 & \rho_{1u} & \rho_{2u} \\ \rho_{1u} & 1 & \rho_{12} \\ \rho_{2u} & \rho_{12} & 1 \end{bmatrix}. \quad (3)$$

The parameter vector $\boldsymbol{\theta}$ for the Roy model is

$$\boldsymbol{\theta} = [\boldsymbol{\gamma}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \rho_{1u}, \rho_{2u}, \rho_{12}]' = [\boldsymbol{\varphi}, \rho_{1u}, \rho_{2u}, \rho_{12}]',$$

where

$$\boldsymbol{\varphi} = [\boldsymbol{\gamma}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2].$$

In the discussion that follows conditioning on $\boldsymbol{\varphi}$ will be suppressed; like Vijverberg (1993) and Koop and Poirier (1997), $\boldsymbol{\varphi}$ will be treated as nuisance parameters.

In constructing the likelihood function for the extended Roy model, two cases are distinguished.

2.1 Full observability

Full observability refers to the scenario where the following are observed for each of the $j = 1, \dots, n$ agents: (a) potential earnings in sectors 1 and 2, Y_1^*, Y_2^* , respectively; and (b) the difference in utility between sector 1 and 2, I^* .

The likelihood function under *full observability* is the product over $j = 1, \dots, n$ individuals of the density function of the trivariate Normal distribution given in equation (2).

$$f_{fo}(\mathbf{z}|\rho_{12}, \rho_{1u}, \rho_{2u}) = (2\pi)^{-1/2} |\boldsymbol{\Omega}|^{-1} e^{-\frac{1}{2}\mathbf{z}'\boldsymbol{\Omega}^{-1}\mathbf{z}}, \quad (4)$$

where $\mathbf{z} = [I^* - \mathbf{W}\boldsymbol{\gamma}, Y_1^* - \mathbf{X}_1\boldsymbol{\beta}_1, Y_2^* - \mathbf{X}_2\boldsymbol{\beta}_2]'$.

2.2 Partial observability

Of the n agents I shall assume that n_1 choose sector 1 and n_2 choose sector 2. Following Koop and Poirier (1997, p. 221), I rewrite the potential earnings in the two sectors as

$$\mathbf{Y}_1^* = \begin{bmatrix} \mathbf{Y}_1^O \\ \mathbf{Y}_1^M \end{bmatrix}, \quad \mathbf{Y}_2^* = \begin{bmatrix} \mathbf{Y}_2^O \\ \mathbf{Y}_2^M \end{bmatrix}, \quad (5)$$

³For example see Roy (1951), Poirier and Rudd (1981), Madalla (1987), Vijverberg (1993) and Koop and Poirier (1997).

where the superscript “ O ” stands for observed data and “ M ” for missing data. \mathbf{Y}_1^O is an n_1 dimensional vector; it indicates the realized earnings of n_1 individuals in sector 1 when they choose sector 1. The potential earnings of these n_1 individuals in sector 2, the sector they did not choose, are given by the n_1 dimensional vector \mathbf{Y}_2^M . \mathbf{Y}_1^M is an n_2 dimensional vector of potential earnings of n_2 individuals in sector 1. The sector of choice of these n_2 individuals is sector 2; their realized earnings in this sector are given by the n_2 dimensional vector \mathbf{Y}_2^O .

Since an individual chooses either sector 1 or sector 2, under *partial observability* all that is *ever* observed is this individual’s realized earnings in the sector of choice; these will be indicated by Y_1^O or Y_2^O depending on whether sector 1 or sector 2 is chosen by individual j . The potential earnings of individual j in the sector that she did *not* choose, are *not* observed.

Thus, under *partial observability*, the following are observed for each agent $j = 1, \dots, n$: (a) realized earnings in the sector of choice, Y_1^O or Y_2^O ; (b) sectoral choice, I , where

$$Y_1^O = \begin{cases} Y_1^* & \text{iff } I^* > 0 \\ 0 & \text{otherwise.} \end{cases} \quad Y_2^O = \begin{cases} Y_2^* & \text{iff } I^* \leq 0 \\ 0 & \text{otherwise.} \end{cases} \quad I = \begin{cases} 1 & \text{iff } I^* > 0 \\ 0 & \text{iff } I^* \leq 0. \end{cases}$$

The likelihood function under *partial observability* is the product over $j = 1, \dots, n$ of the following density function,

$$f_{po}(y_1^O, y_2^O, I | \rho_{12}, \rho_{1u}, \rho_{2u}) = \left[\int_0^\infty f_{1u}(y_1^O, u | \rho_{1u}, \rho_{2u}) du \right]^I \left[\int_{-\infty}^0 f_{2u}(y_2^O, u | \rho_{1u}, \rho_{2u}) du \right]^{1-I}. \quad (6)$$

$f_{1u}(\bullet | \boldsymbol{\varphi}, \rho_{1u}, \rho_{2u})$ and $f_{2u}(\bullet | \boldsymbol{\varphi}, \rho_{1u}, \rho_{2u})$ are density functions of bivariate normal distributions. These can be obtained from the trivariate Normal distribution specified in (2) (Poirier, 1995, p. 122).⁴

The *key* difference between the *partial observability* and the *full observability* likelihood function is that the correlation coefficient between unobserved productivity in the two sectors, ρ_{12} , does not appear in the former likelihood function given in equation (6). Thus, ρ_{12} is not identified under *partial observability*. The intuition for this is simple. Under *partial observability* earnings of an agent in the sector of choice are observed; potential earnings in the sector that she did not choose are *not* observed. For ρ_{12} , the correlation coefficient of potential earnings between unobserved productivity in the two sectors to be identified, the earnings of some individuals in *both* sectors have to be observed.

⁴Thus, for $i = 1, 2$, $[Y_i^O, U]'$ follows a bivariate Normal distribution with mean $[\mathbf{X}_i \boldsymbol{\beta}_1, \mathbf{W} \boldsymbol{\gamma}]'$ and variance-covariance matrix

$$\begin{bmatrix} 1 & \rho_{iu} \\ \rho_{iu} & 1 \end{bmatrix}.$$

Since under *full observability*, an agent's earnings in the sector she chose, as well as, the sector she did not choose, are observed, ρ_{12} is identified.

But Vijverberg (1993, p. 71) points out that data contains expected information about ρ_{12} under *partial observability*. The positive definiteness of Ω limits the range of ρ_{12} when ρ_{1u} and ρ_{2u} are freely chosen from the interval $[-1, 1]$. This restricted support of ρ_{12} will be indicated by $\Psi(\rho_{1u}, \rho_{2u})$, where

$$\Psi(\rho_{1u}, \rho_{2u}) = [\rho_{12}^l, \rho_{12}^u] = \left\{ \rho_{12} \mid \rho_{12} \in \left[\rho_{1u}\rho_{2u} \mp \sqrt{(1 - \rho_{1u}^2)(1 - \rho_{2u}^2)} \right] \right\}; \quad (7)$$

ρ_{12}^l, ρ_{12}^u are the lower and upper bounds of this support, respectively. Hence under *partial observability* data contains expected information about ρ_{12} through the parameters ρ_{1u} and ρ_{2u} which are identified from the observed data on realized earnings and sectoral choice.

This motivates rewriting the likelihood function under *partial observability* given in equation (6) as

$$f_{po}(y_1^O, y_2^O, I \mid \rho_{12}, \rho_{1u}, \rho_{2u}) = \mathbf{1}_{\Psi(\rho_{1u}, \rho_{2u})}(\rho_{12}), \quad (8)$$

where $\mathbf{1}_{\Psi(\rho_{1u}, \rho_{2u})}(\rho_{12})$ is an indicator function with

$$\begin{aligned} \mathbf{1}_{\Psi(\rho_{1u}, \rho_{2u})}(\rho_{12}) &= 1 && \text{if } \rho_{12} \in \Psi(\rho_{1u}, \rho_{2u}) \\ &= 0 && \text{otherwise.} \end{aligned} \quad (9)$$

3 Prior and Posterior Analysis for $\rho_{12} \mid \rho_{1u}, \rho_{2u}$

Typically, in applied work on the extended Roy model, a researcher has access to data corresponding to the *partial observability* case. As pointed out in section 2, ρ_{12} does not enter the likelihood function explicitly under *partial observability*. But the positive definiteness of Ω bounds ρ_{12} ; equation (7) gives the interval within which ρ_{12} is restricted by the positive definiteness of Ω . Vijverberg (1993) and Koop and Poirier (1997) use this to specify a uniform prior for $\rho_{12} \mid \rho_{1u}, \rho_{2u}$;

$$\pi^{po}(\rho_{12} \mid \rho_{1u}, \rho_{2u}) = \frac{1}{\rho_{12}^u - \rho_{12}^l}, \quad \rho_{12}^l \leq \rho_{12} \leq \rho_{12}^u. \quad (10)$$

Using this prior, Koop and Poirier (1997) show that under *partial observability*, while the marginal prior and posterior for ρ_{12} are different, the prior and posterior for ρ_{12} conditional on ρ_{1u}, ρ_{2u} , are identical. Thus, under *partial observability*, data modifies prior beliefs about ρ_{12} but not about $\rho_{12} \mid \rho_{1u}, \rho_{2u}$.

The intuition for this result comes from a more general setting: Bayesian inference of nonidentified models (Dawid, 1979 and Poirier, 1998). As long as *proper* priors are specified on all parameters,⁵ prior beliefs about a nonidentified

⁵A distribution is proper if its density function integrates to 1. It is improper if the integral of the density function diverges to infinity.

parameter are modified by data through the identified parameters, if the two groups of parameters are apriori dependent. Since the identified parameters are the “vehicle” of learning for the nonidentified parameter, the marginal prior and posterior for the nonidentified parameter differ, but the prior and posterior for the nonidentified parameter, *conditional* on the identified parameters, are identical. Thus, marginal inference about the unidentified parameter is feasible, but conditional inference is not. Since ρ_{1u}, ρ_{2u} are identified and ρ_{12} nonidentified under *partial observability*, the intuition of the literature on nonidentified models carries over to the extended Roy model of selectivity.

This intuition *does not* generalize to the case when *improper* priors are specified for the nonidentified parameters. Poirier (1998) and Ghosh, et al (1999) show through several examples that when a nonidentified parameter is combined with an improper prior for the nonidentified parameter, whether the posterior is proper depends on the nature of nonidentifiability and the *kind of information* provided by the prior even if it is improper.

This Section shows that under *partial observability*, if Jeffreys’ prior for ρ_{12} based on the sampling density under *full observability* is used, **data modifies prior beliefs about the unidentified parameter, ρ_{12} , conditional on the identified parameters, ρ_{1u} and ρ_{2u}** . Hence, inference about $\rho_{12}|\rho_{1u}, \rho_{2u}$ becomes feasible when the data observed by a researcher corresponds to the *partial observability* case.

3.1 Jeffreys’ Prior under *Full Observability*

Jeffreys’ prior for $\rho_{12}|\rho_{1u}, \rho_{2u}$ under *full observability* or *partial observability* is the square-root of the Fisher information about ρ_{12} under that scenario. Since ρ_{12} does not enter the likelihood function under *partial observability*, the Fisher information about ρ_{12} is undefined under *partial observability*.

Given ρ_{1u}, ρ_{2u} , the Fisher information about ρ_{12} under *full observability* has been obtained in Appendix A. The square-root of this Fisher information is the Jeffreys’ prior for ρ_{12} conditional on ρ_{1u}, ρ_{2u} . It is given by

$$\pi^{fo}(\rho_{12}|\rho_{1u}, \rho_{2u}) \propto \sqrt{\frac{1}{|\mathbf{\Omega}|} \left[1 + \frac{2}{|\mathbf{\Omega}|} (\rho_{1u}\rho_{2u} - \rho_{12})^2 \right]}, \quad -1 \leq \rho_{12} \leq 1; \quad (11)$$

$|\mathbf{\Omega}|$ is the determinant of the covariance matrix $\mathbf{\Omega}$ given in equation (3),

$$|\mathbf{\Omega}| = 1 - \rho_{1u}^2 - \rho_{2u}^2 - \rho_{12}^2 + 2\rho_{1u}\rho_{2u}\rho_{12}. \quad (12)$$

Jeffreys’ prior given in the above equation is improper or diverges to infinity. The prior puts substantial mass at the end-points of the interval $\Psi(\rho_{1u}, \rho_{2u})$ given by equation (7). That is,

$$\pi^{fo}(\rho_{12}|\rho_{1u}, \rho_{2u}) \rightarrow \infty \quad \text{as} \quad \rho_{12} \rightarrow \rho_{12}^l, \rho_{12}^u.$$

This follows from ρ_{12}^l, ρ_{12}^u being solutions to the equation $|\mathbf{\Omega}| = 0$.⁶ Hence, $|\mathbf{\Omega}| \rightarrow 0$ as $\rho_{12} \rightarrow \rho_{12}^l, \rho_{12}^u$.

In addition, **conditional** on values of ρ_{1u}, ρ_{2u} , $\pi^{fo}(\rho_{12}|\rho_{1u}, \rho_{2u}) \rightarrow \infty$ as $\rho_{12} \rightarrow +1, -1$. This follows from $|\mathbf{\Omega}| \rightarrow -(\rho_{1u} - \rho_{2u})^2$ as $\rho_{12} \rightarrow +1$, and $|\mathbf{\Omega}| \rightarrow -(\rho_{1u} + \rho_{2u})^2$ as $\rho_{12} \rightarrow -1$. Thus, $\pi^{fo}(\rho_{12}|\rho_{1u}, \rho_{2u}) \rightarrow \infty$ if $\rho_{1u} = \rho_{2u}$ in the former case, and $\rho_{1u} = -\rho_{2u}$ in the latter case.

The **key** point that emerges is that Jeffreys' prior is not well-behaved outside the interval $\Psi(\rho_{1u}, \rho_{2u})$ given by equation (7).

3.2 Posterior for $\rho_{12}|\rho_{1u}, \rho_{2u}$

When observed data corresponds to the *partial observability* case, the likelihood function is given by equation (8). Using Jeffreys' prior obtained in equation (11), the posterior for $\rho_{12}|\rho_{1u}, \rho_{2u}$ is

$$\pi(\rho_{12}|\rho_{1u}, \rho_{2u}, \mathbf{y}_1^O, \mathbf{y}_2^O, \mathbf{I}) \propto \pi^{fo}(\rho_{12}|\rho_{1u}, \rho_{2u}) \mathbf{1}_{\Psi(\rho_{1u}, \rho_{2u})}(\rho_{12}). \quad (13)$$

The posterior for $\rho_{12}|\rho_{1u}, \rho_{2u}$ is Jeffreys' prior for ρ_{12} under *full observability*, **restricted** to the interval $\Psi(\rho_{1u}, \rho_{2u}) = [\rho_{12}^l, \rho_{12}^u]$ by the *partial observability* likelihood function. The posterior for $\rho_{12}|\rho_{1u}, \rho_{2u}$ given by equation (13) has been plotted in Figure 1. Two data sets used by Vijverberg (1993, pp. 80-83) have been used to generate the posterior distribution. The estimates of (ρ_{1u}, ρ_{2u}) differ in the two data sets. As a result, the support of the posterior distribution is different in the two cases even though the shape of the posterior is similar.

Section 3.1 shows that Jeffreys' prior is not well-behaved outside the interval $\Psi(\rho_{1u}, \rho_{2u})$. Specifically, $\pi^{fo}(\rho_{12}|\rho_{1u}, \rho_{2u}) \rightarrow \infty$ as $\rho_{12} \rightarrow \rho_{12}^l, \rho_{12}^u$, and $\rho_{12} \rightarrow +1, -1$. The posterior for $\rho_{12}|\rho_{1u}, \rho_{2u}$ would be proper if it was possible to restrict the prior to the interval $\Psi(\rho_{1u}, \rho_{2u})$. This is what the likelihood function accomplishes. The likelihood function restricts $\mathbf{\Omega}$ to be positive definite; a necessary condition for this is $|\mathbf{\Omega}| > 0$. Equivalently, in terms of ρ_{12} , $\rho_{12}^l < \rho_{12} < \rho_{12}^u$. A formal proof that the posterior is proper is provided in Appendix C.

Reflecting properties of the prior, the posterior distribution has three important features.

First, it is approximately symmetric around the mid-point, ρ_{12}^m , of the interval $\Psi(\rho_{1u}, \rho_{2u})$. From equation (7), the mid-point of $\Psi(\rho_{1u}, \rho_{2u})$ is

$$\rho_{12}^m = \rho_{1u}\rho_{2u}.$$

Symmetry is proved by showing that

$$Pr(\rho_{12}^l \leq \rho_{12} \leq \rho_{12}^m) = Pr(\rho_{12}^m \leq \rho_{12} \leq \rho_{12}^u) = 0.5.$$

⁶From equation (12), $|\mathbf{\Omega}|$ is a quadratic function of ρ_{12} .

Normalizing the posterior kernel in equation (13) so that it integrates to 1, this is equivalent to showing that

$$\int_{\rho_{12}^l}^{\rho_{12}^m} \pi(\rho_{12} | \rho_{1u}, \rho_{2u}, \mathbf{y}_1^O, \mathbf{y}_2^O, \mathbf{I}) d\rho_{12} = \int_{\rho_{12}^m}^{\rho_{12}^u} \pi(\rho_{12} | \rho_{1u}, \rho_{2u}, \mathbf{y}_1^O, \mathbf{y}_2^O, \mathbf{I}) d\rho_{12} = 0.5.$$

Numerical integration is used to obtain the two integrals and verify that they have a value of 0.5. This exercise was repeated for several values of (ρ_{1u}, ρ_{2u}) in the interval $[-0.99, +0.99]$ including the estimates obtained for the two data sets used by Vijverberg (1993, pp. 80-83).

Second, it has two points of inflection at

$$\rho_{1u}\rho_{2u} \pm 0.65\sqrt{(1 - \rho_{1u}^2)(1 - \rho_{2u}^2)}.$$

These have been obtained in Appendix B.

Third, the the posterior has two modes at the end-points of the interval $\Psi(\rho_{1u}, \rho_{2u})$ given by equation (7).

Thus, when observed data corresponds to the *partial observability* case, inference about ρ_{12} conditional on the identified parameters ρ_{1u}, ρ_{2u} is feasible, if Jeffreys' prior under *full observability* is used.

4 DISCUSSION

Several interesting issues arise at this stage.

First, the fact that *both* observed and missing data provide **more** expected information about parameters than just the observed data, is not new. It was first proposed as the “Missing Information Principle” by Woodbury (1977, pp. 808-813). Treating missing data as random variables, the “Missing Information Principle” stated that the total information about parameters is the sum of the information contained in the available data plus the lost information in the missing data. From Woodbury (1977, p. 809),

$$\mathbf{J}_{\mathbf{Y}^O, \mathbf{Y}^M}(\boldsymbol{\theta}) = \mathbf{J}_{\mathbf{Y}^O}(\boldsymbol{\theta}) + E_{\mathbf{Y}^O}[\mathbf{J}_{\mathbf{Y}^O, \mathbf{Y}^M | \mathbf{Y}^O}(\boldsymbol{\theta})] \quad (14)$$

$\mathbf{Y}^O, \mathbf{Y}^M$ is the observed and missing data, respectively. The left-hand-side is the Fisher information about $\boldsymbol{\theta}$ provided by the observed and missing data. The first term on the right-hand-side is the Fisher information about $\boldsymbol{\theta}$ provided by the observed data. The second term is the lost information in the missing data. It is the expectation of the Fisher information about $\boldsymbol{\theta}$ provided by the missing data conditional on the observed data; the expectation is with respect to the sampling density of the observed data.⁷ An exception to the “Missing

⁷This principle has been extended to the case of nonregular likelihood functions by Sareen (2000, pp. 47-51). A likelihood function is nonregular if the support of the data depends on parameters.

Information Principle” that is well recognized is when the observed data is *jointly sufficient* for θ . Thus,

$$\mathbf{J}_{\mathbf{Y}^O, \mathbf{Y}^M}(\theta) = \mathbf{J}_{\mathbf{Y}^O}(\theta)$$

if and only if \mathbf{Y}^O is jointly sufficient for θ .

The intuition of the “Missing Information Principle” has been applied extensively in the literature on estimation of missing data problems using the EM algorithm; Meng and van Dyk, 1997 provide an exhaustive survey. Specifically, the speed at which the algorithm converges follows from the “Missing Information Principle”: the fraction of missing information determines the speed of convergence.

Since Jeffreys’ prior is the square-root of the Fisher information, this paper establishes a link, through the Jeffreys’ prior, between the “Missing Information Principle” and Bayesian estimation of an unidentified parameter in a *missing data* problem. Specifically, in the Roy model, Jeffreys’ prior for $\rho_{12}|\rho_{1u}, \rho_{2u}$ under *full observability* provides more expected information about $\rho_{12}|\rho_{1u}, \rho_{2u}$ than Jeffreys’ prior for $\rho_{12}|\rho_{1u}, \rho_{2u}$ under *partial observability*. This *extra* prior information makes inference about $\rho_{12}|\rho_{1u}, \rho_{2u}$ feasible.

This leads to the *second* issue. What *extra* information is provided by Jeffrey’s prior based on the *full observability* likelihood that is not provided by Jeffrey’s prior based on the *partial observability* likelihood? The “Missing Information Principle”, stated in equation (14), cannot be used directly to quantify this *extra* information. This follows from the Fisher information for ρ_{12} , and Jeffreys’ prior, being undefined under *partial observability* as the likelihood is an indicator function. The uniform prior for $\rho_{12}|\rho_{1u}, \rho_{2u}$ in equation (10), used by Vijverberg (1993) and Koop and Poirier (1997), captures the expected information provided by data under *partial observability* by restricting the support of ρ_{12} as a function of the identified parameters, ρ_{1u}, ρ_{2u} . Hence, I shall compare Jeffreys’ prior for $\rho_{12}|\rho_{1u}, \rho_{2u}$ under *full observability* with the uniform prior for $\rho_{12}|\rho_{1u}, \rho_{2u}$ given in equation (10) to understand what *extra* information is provided by Jeffrey’s prior based on the *full observability* likelihood.

The Roy model has empirical content if $\rho_{12} \neq 0$. The three features of Jeffreys’ prior under *full observability* described in Section 3.2, ensure this. Suppose a configuration of ρ_{1u}, ρ_{2u} is such that the mid-point of the interval $\Psi(\rho_{1u}, \rho_{2u})$ is at 0; then Jeffreys’ prior avoids this part of the parameter space of ρ_{12} since most of its mass is at the tails. It is conceivable that a configuration of ρ_{1u}, ρ_{2u} leads to one of the end-points of the interval $\Psi(\rho_{1u}, \rho_{2u})$ being 0. By distributing mass equally on either side of the mid-point of the interval $\Psi(\rho_{1u}, \rho_{2u})$, it makes a part of the parameter space of ρ_{12} , quite different from 0, equally likely, restoring the empirical content of the Roy model. This is in contrast to the uniform prior for $\rho_{12}|\rho_{1u}, \rho_{2u}$ in equation (10). This prior makes each point on the interval $\Psi(\rho_{1u}, \rho_{2u})$ *equally likely*.

Third, Koop and Poirier (1997, p. 219) state that the prior and posterior for $\rho_{12}|\rho_{1u}, \rho_{2u}$ are identical, when the uniform prior is combined with the *partial observability* likelihood. The inability of data under *partial observability*, to update beliefs about $\rho_{12}|\rho_{1u}, \rho_{2u}$, is attributed to the conditioning on the identified parameters. It appears that it is not just the conditioning on the identified parameters, but the *manner* in which the *conditioning* is being done, that is driving the prior and the posterior for $\rho_{12}|\rho_{1u}, \rho_{2u}$, to be identical. This becomes obvious if I write out the the posterior for $\rho_{12}|\rho_{1u}, \rho_{2u}$ based on the uniform prior given in equation (10), and the *partial observability* likelihood given by equation (8);

$$\pi(\rho_{12}|\rho_{1u}, \rho_{2u}, \mathbf{y}_1^O, \mathbf{y}_2^O, \mathbf{I}) \propto \left[\frac{1}{\rho_{12}^u - \rho_{12}^l} \right] \mathbf{1}_{\Psi(\rho_{1u}, \rho_{2u})}(\rho_{12}). \quad (15)$$

Both the likelihood function and the prior appear to use the same *channel* to extract information about the nonidentified parameter from the identified parameters: the positive definiteness of $\mathbf{\Omega}$. Since the likelihood and the prior are “identical” in this sense, it is not surprising that prior beliefs about $\rho_{12}|\rho_{1u}, \rho_{2u}$ do not get updated.

Fourth, while this paper deals only with the Roy model, missing data is encountered in several popular applications in economics. In view of the “Missing Information Principle”, in case noninformative priors are being specified, an empirical researcher may consider using a noninformative prior that is based on the likelihood of the complete data rather than the observed, incomplete data. While results as strong as the Roy model may not emerge, both point estimates and interval estimates are likely to be more accurate. In addition, sampling from the posterior could be less cumbersome since in most cases the complete data likelihood, and hence the prior based on it, has a simpler form than the incomplete data likelihood and the prior based on it.

ACKNOWLEDGEMENT

I have benefited from invaluable discussions with Malay Ghosh, Tony Lancaster, Mike Peters, Dale J. Poirier, and participants at the International Workshop on Objective Bayesian methodology, Valência, 1999 and Sixth World meeting of the ISBA, Greece, 2000. Comments from three reviewers of a previous paper are gratefully acknowledged.

APPENDIX

A Fisher Information under *Full Observability*

The log likelihood function for the extended Roy model under *full observability* is

$$\log f_{fo}(\mathbf{z}|\rho_{12}, \rho_{1u}, \rho_{2u}) = -\frac{1}{2} \log |\mathbf{\Omega}| - \frac{1}{2} \mathbf{z}' \mathbf{\Omega}^{-1} \mathbf{z}, \quad (\text{A.1})$$

where $\mathbf{z} = [z_I, z_1, z_2]' = [I^* - \mathbf{W}\boldsymbol{\gamma}, Y_1^* - \mathbf{X}_1\boldsymbol{\beta}_1, Y_2^* - \mathbf{X}_2\boldsymbol{\beta}_2]'$. $|\boldsymbol{\Omega}|$ is the determinant of the covariance matrix $\boldsymbol{\Omega}$; it is restricted to be *strictly* greater than zero from the positive definiteness of $\boldsymbol{\Omega}$. I shall indicate it by D ,

$$D = |\boldsymbol{\Omega}| = 1 - \rho_{1u}^2 - \rho_{2u}^2 - \rho_{12}^2 + 2\rho_{1u}\rho_{2u}\rho_{12}. \quad (\text{A.2})$$

Simplifying $\mathbf{z}'\boldsymbol{\Omega}^{-1}\mathbf{z}$, I obtain

$$\mathbf{z}'\boldsymbol{\Omega}^{-1}\mathbf{z} = \frac{A}{D}, \quad (\text{A.3})$$

where

$$\begin{aligned} A = & (1 - \rho_{12}^2)z_I^2 + (1 - \rho_{2u}^2)z_1^2 + (1 - \rho_{1u}^2)z_2^2 + \\ & 2(\rho_{12}\rho_{2u} - \rho_{1u})z_I z_1 + 2(\rho_{12}\rho_{1u} - \rho_{2u})z_I z_2 + \\ & 2(\rho_{1u}\rho_{2u} - \rho_{12})z_1 z_2. \end{aligned} \quad (\text{A.4})$$

I can rewrite the loglikelihood function in (A.1) as

$$\log f_{fo}(\mathbf{z}|\rho_{12}, \rho_{1u}, \rho_{2u}) = -\frac{1}{2} \log D - \frac{1}{2} \frac{A}{D}. \quad (\text{A.5})$$

The score of this loglikelihood function with respect to ρ_{12} is

$$S_{fo}^{\rho_{12}}(\rho_{12}, \rho_{1u}, \rho_{2u}) = \frac{d \log f_{I^* Y_1^* Y_2^*}(\mathbf{z}|\rho_{12}, \rho_{1u}, \rho_{2u})}{d\rho_{12}} = -\frac{1}{2} \left[\frac{1}{D} \frac{dD}{d\rho_{12}} + \frac{1}{D^2} \left(D \frac{dA}{d\rho_{12}} - A \frac{dD}{d\rho_{12}} \right) \right];$$

taking the derivative of the score with respect to ρ_{12} , the Hessian is

$$\begin{aligned} H_{fo}^{\rho_{12}}(\rho_{12}, \rho_{1u}, \rho_{2u}) = & -\frac{1}{2D} \left[\frac{d^2 D}{d\rho_{12}^2} \left(1 - \frac{A}{D} \right) + \frac{d^2 A}{d\rho_{12}^2} + \frac{1}{D} \left(\frac{dD}{d\rho_{12}} \right)^2 \left(\frac{2A}{D} - 1 \right) - \frac{2}{D} \frac{dA}{d\rho_{12}} \frac{dD}{d\rho_{12}} \right]; \\ = & -\frac{1}{D} \left[-\left(1 - \frac{A}{D} \right) - Z_I^2 + \frac{2}{D} (\rho_{1u}\rho_{2u} - \rho_{12})^2 \left(\frac{2A}{D} - 1 \right) \right. \\ & \left. - \frac{4}{D} (\rho_{1u}\rho_{2u} - \rho_{12}) (\rho_{2u}Z_I Z_1 + \rho_{1u}Z_I Z_2 - \rho_{12}Z_I^2 - Z_1 Z_2) \right]. \end{aligned} \quad (\text{A.6})$$

The last step in (A.6) has been obtained by substituting the following derivatives in the previous step.

$$\begin{aligned} \frac{dD}{d\rho_{12}} = & 2[\rho_{1u}\rho_{2u} - \rho_{12}] \quad \frac{d^2 D}{d\rho_{12}^2} = -2 \quad \text{and}; \\ \frac{dA}{d\rho_{12}} = & 2[-\rho_{12}Z_I^2 + \rho_{2u}Z_I Z_1 + \rho_{1u}Z_I Z_2 - Z_1 Z_2] \quad \frac{d^2 A}{d\rho_{12}^2} = -2Z_I^2. \end{aligned}$$

The Fisher information about ρ_{12} under *full observability* is the expected value of the Hessian in equation (A.6). The expectation is with respect to the sampling density under *full observability* given in equation (2),

$$\mathbf{Z} = [Z_I, Z_1, Z_2]' \sim N_3(\mathbf{0}, \boldsymbol{\Omega}) \quad \text{with} \quad \boldsymbol{\Omega} = \begin{bmatrix} 1 & \rho_{1u} & \rho_{2u} \\ \rho_{1u} & 1 & \rho_{12} \\ \rho_{2u} & \rho_{12} & 1 \end{bmatrix}. \quad (\text{A.7})$$

Thus, the Fisher information about ρ_{12} under *full observability*, $J_{f_o}^{\rho_{12}}(\rho_{12}, \rho_{1u}, \rho_{2u})$, is

$$\begin{aligned}
J_{f_o}^{\rho_{12}}(\rho_{12}, \rho_{1u}, \rho_{2u}) &= -E_{\mathbf{z}} \left[H_{f_o}^{\rho_{12}}(\rho_{12}, \rho_{1u}, \rho_{2u}) \right] \\
&= \frac{1}{D} \left[- \left(1 - \frac{E_{\mathbf{z}}(A)}{D} \right) - E_{z_I}(Z_I^2) + \frac{2}{D} (\rho_{1u}\rho_{2u} - \rho_{12})^2 \left(\frac{2E_{\mathbf{z}}(A)}{D} - 1 \right) \right. \\
&\quad - \frac{4}{D} (\rho_{1u}\rho_{2u} - \rho_{12})\rho_{2u} E_{z_I z_1}(Z_{Ij} Z_{1j}) - \frac{4}{D} (\rho_{1u}\rho_{2u} - \rho_{12})\rho_{1u} E_{z_I z_2}(Z_{Ij} Z_{2j}) \\
&\quad \left. + \frac{4}{D} (\rho_{1u}\rho_{2u} - \rho_{12})\rho_{12} E_{z_I}(Z_{Ij}^2) + \frac{4}{D} (\rho_{1u}\rho_{2u} - \rho_{12}) E_{z_1 z_2}(Z_{1j} Z_{2j}) \right]. \tag{A.8}
\end{aligned}$$

From the trivariate Normal distribution given by (A.7), the bivariate and marginal distributions can be obtained (Poirier, 1995, pp. 121-124). Thus,

$$E_{\mathbf{z}}(Z_I^2) = E_{\mathbf{z}}(Z_1^2) = E_{\mathbf{z}}(Z_2^2) = 1; \quad \text{and}$$

$$E_{\mathbf{z}}(Z_I Z_1) = \rho_{1u}, \quad E_{\mathbf{z}}(Z_I Z_2) = \rho_{2u}, \quad \text{and } E_{\mathbf{z}}(Z_1 Z_2) = \rho_{12}.$$

Consequently $E_{\mathbf{z}}(A) = 3D$ can be obtained from (A.4). Substituting these expected values in (A.8),

$$J_{f_o}^{\rho_{12}}(\rho_{12}, \rho_{1u}, \rho_{2u}) = \frac{1}{|\mathbf{\Omega}|} \left[1 + \frac{2}{|\mathbf{\Omega}|} (\rho_{1u}\rho_{2u} - \rho_{12})^2 \right].$$

Taking the square-root of this, gives the Jeffreys' prior for $\rho_{12} | \rho_{1u}, \rho_{2u}$ under *full observability*.

B Points of Inflexion of Jeffreys' Prior under *Full Observability*

From Section 3.1, equation (11), Jeffreys' prior under *full Observability* is

$$\pi^{f_o}(\rho_{12} | \rho_{1u}, \rho_{2u}) \propto \sqrt{\frac{1}{D} \left[1 + \frac{2}{D} (\rho_{1u}\rho_{2u} - \rho_{12})^2 \right]}, \quad -1 \leq \rho_{12} \leq 1; \tag{B.1}$$

where D is the determinant of the covariance matrix, $\mathbf{\Omega}$,

$$D = |\mathbf{\Omega}| = 1 - \rho_{1u}^2 - \rho_{2u}^2 - \rho_{12}^2 + 2\rho_{1u}\rho_{2u}\rho_{12}. \tag{B.2}$$

To obtain the points of inflexion, I need to obtain the first derivative of the prior in (B.1) with respect to ρ_{12} ,

$$\frac{d\pi^{f_o}(\rho_{12} | \rho_{1u}, \rho_{2u})}{d\rho_{12}} = \frac{\frac{d}{d\rho_{12}} \left[\frac{1}{D} + \frac{2}{D^2} (\rho_{1u}\rho_{2u} - \rho_{12})^2 \right]}{2\sqrt{\frac{1}{D} \left[1 + \frac{2}{D} (\rho_{1u}\rho_{2u} - \rho_{12})^2 \right]}}. \tag{B.3}$$

Evaluating the numerator in the above equation,

$$\begin{aligned}
\frac{d}{d\rho_{12}} \left[\frac{1}{D} + \frac{2}{D^2} (\rho_{1u}\rho_{2u} - \rho_{12})^2 \right] &= \frac{d}{d\rho_{12}} \left[\frac{1}{D} \right] + 2 \frac{d}{d\rho_{12}} \left[\frac{1}{D^2} (\rho_{1u}\rho_{2u} - \rho_{12})^2 \right], \\
&= \left[-\frac{1}{D^2} \frac{dD}{d\rho_{12}} \right] + 2 \left[-\frac{2}{D^2} (\rho_{1u}\rho_{2u} - \rho_{12}) - \frac{2}{D^3} (\rho_{1u}\rho_{2u} - \rho_{12})^2 \frac{dD}{d\rho_{12}} \right], \\
&= -\frac{4}{D^2} (\rho_{1u}\rho_{2u} - \rho_{12}) - 2(\rho_{1u}\rho_{2u} - \rho_{12}) \left[\frac{1}{D^2} + \frac{4}{D^3} (\rho_{1u}\rho_{2u} - \rho_{12})^2 \right].
\end{aligned} \tag{B.4}$$

The last step in (B.4) follows from substituting

$$\frac{dD}{d\rho_{12}} = 2(\rho_{1u}\rho_{2u} - \rho_{12}),$$

and simplifying the resulting expression.

The points of inflexion of Jeffreys' prior under *full Observability* can be obtained by equating the derivative in (B.4) to zero;

$$\frac{d\pi^{fo}(\rho_{12}|\rho_{1u}, \rho_{2u})}{d\rho_{12}} = 0 \Leftrightarrow 7\rho_{12}^2 - 14\rho_{1u}\rho_{2u}\rho_{12} + C = 0,$$

where $C = 4\rho_{1u}^2\rho_{2u}^2 + 3\rho_{1u}^2 + 3\rho_{2u}^2 - 3$. Solving the above quadratic for ρ_{12} obtains the points of inflexion given in Section 3.1

C Proving the Posterior for $\rho_{12}|\rho_{1u}, \rho_{2u}$ is Integrable

From equations (11) and (13), the posterior for $\rho_{12}|\rho_{1u}, \rho_{2u}$, when the observed data corresponds to the *partial observability* scenario, is

$$\pi(\rho_{12}|\rho_{1u}, \rho_{2u}, \mathbf{y}_1^O, \mathbf{y}_2^O, \mathbf{I}) \propto \sqrt{\frac{1}{|\boldsymbol{\Omega}|} \left[1 + \frac{2}{|\boldsymbol{\Omega}|} (\rho_{1u}\rho_{2u} - \rho_{12})^2 \right]} \mathbf{1}_{\Psi(\rho_{1u}, \rho_{2u})}(\rho_{12}).$$

To prove that this posterior is proper, I shall have to show that

$$\int_{\rho_{12}^l}^{\rho_{12}^u} \sqrt{\left[\frac{1}{D} + \frac{2}{D^2} (\rho_{1u}\rho_{2u} - \rho_{12})^2 \right]} d\rho_{12}, \tag{C.1}$$

where $D = |\boldsymbol{\Omega}|$,

$$\rho_{12}^l = \rho_{1u}\rho_{2u} - \sqrt{(1 - \rho_{1u}^2)(1 - \rho_{2u}^2)}, \quad \text{and } \rho_{12}^u = \rho_{1u}\rho_{2u} + \sqrt{(1 - \rho_{1u}^2)(1 - \rho_{2u}^2)}. \tag{C.2}$$

Since $D = |\mathbf{\Omega}|$, $D > 0$ follows from the covariance matrix $\mathbf{\Omega}$ being positive definite. Hence,

$$\begin{aligned} \int_{\rho_{12}^l}^{\rho_{12}^u} \sqrt{\left[\frac{1}{D} + \frac{2(\rho_{1u}\rho_{2u} - \rho_{12})^2}{D^2} \right]} d\rho_{12} &\leq \int_{\rho_{12}^l}^{\rho_{12}^u} \left[\sqrt{\frac{1}{D}} + \sqrt{\frac{2(\rho_{1u}\rho_{2u} - \rho_{12})^2}{D^2}} \right] d\rho_{12}, \\ &= \int_{\rho_{12}^l}^{\rho_{12}^u} \sqrt{\frac{1}{D}} d\rho_{12} + \sqrt{2} \int_{\rho_{12}^l}^{\rho_{12}^u} \frac{(\rho_{1u}\rho_{2u} - \rho_{12})}{D} d\rho_{12}. \end{aligned} \quad (\text{C.3})$$

The inequality I shall now show that each term on the right-hand side of (C.3) is bounded. This will prove that the posterior is integrable.

Starting with the first integral in (C.3),

$$\int_{\rho_{12}^l}^{\rho_{12}^u} \sqrt{\frac{1}{D}} d\rho_{12} = \lim_{\epsilon^l \rightarrow \rho_{12}^l} \lim_{\epsilon^u \rightarrow \rho_{12}^u} \int_{\epsilon^l}^{\epsilon^u} \sqrt{\frac{1}{D}} d\rho_{12}. \quad (\text{C.4})$$

From Gradshteyn and Ryzhik (2000, p. 93),

$$\begin{aligned} \int_{\epsilon^l}^{\epsilon^u} \sqrt{\frac{1}{D}} d\rho_{12} &= -\arcsin \left(\frac{(\rho_{1u}\rho_{2u} - \rho_{12})}{\sqrt{(1 - \rho_{1u}^2)(1 - \rho_{2u}^2)}} \right) \Big|_{\epsilon^l}^{\epsilon^u}, \\ &= -\arcsin \left(\frac{(\rho_{1u}\rho_{2u} - \epsilon^u)}{\sqrt{(1 - \rho_{1u}^2)(1 - \rho_{2u}^2)}} \right) + \arcsin \left(\frac{(\rho_{1u}\rho_{2u} - \epsilon^l)}{\sqrt{(1 - \rho_{1u}^2)(1 - \rho_{2u}^2)}} \right). \end{aligned}$$

Then,

$$\begin{aligned} &\frac{2}{1u}(1 - \rho_{2u}^2) \\ &+ \arcsin \left(\frac{(\rho_{1u}\rho_{2u} - \rho_{12}^l)}{\sqrt{(1 - \rho_{1u}^2)(1 - \rho_{2u}^2)}} \right), \\ &= -\arcsin(-1) + \arcsin(1); \\ &= -\left(\frac{\pi}{2}\right) + \left(\frac{\pi}{2}\right); \\ &= 3.14. \end{aligned}$$

(C.5) follows from substituting ρ_{12}^l, ρ_{12}^u from (C.2).

The second integral in (C.3) is now evaluated.

$$\begin{aligned} \sqrt{2} \int_{\rho_{12}^l}^{\rho_{12}^u} \frac{(\rho_{1u}\rho_{2u} - \rho_{12})}{D} d\rho_{12} &= \frac{1}{\sqrt{2}} \log D \Big|_{\rho_{12}^l}^{\rho_{12}^u} \\ &< \log D \Big|_{\rho_{12}^l}^{\rho_{12}^u}. \end{aligned} \quad (\text{C.7})$$

The indefinite integral is simply the logarithmic function $\log D$. If ρ_{12} is such that, $D \geq 1$, then the indefinite integral is bounded from the properties of

logarithmic functions. However as $D \rightarrow 0$, $\log D \rightarrow -\infty$. But $D > 0$ follows from the covariance matrix $\mathbf{\Omega}$ being positive definite. Intuitively, this is what bounds the integral. A formal proof follows. From (C.7),

$$\begin{aligned} \sqrt{2} \int_{\rho_{12}^l}^{\rho_{12}^u} \frac{(\rho_{1u}\rho_{2u} - \rho_{12})}{D} d\rho_{12} &< \log D \Big|_{\rho_{12}^l}^{\rho_{12}^u}; \\ &\leq e^{\log D} \Big|_{\rho_{12}^l}^{\rho_{12}^u}, \\ &\text{taking a monotonic increasing transformation of } \log D; \\ &= D \Big|_{\rho_{12}^l}^{\rho_{12}^u}; \\ &\leq D \Big|_{-1}^{+1}, \tag{C.8} \\ &= 4\rho_{1u}\rho_{2u}. \tag{C.9} \end{aligned}$$

(C.8) follows from D being a quadratic function of ρ_{12} ,

$$D = |\mathbf{\Omega}| = 1 - \rho_{1u}^2 - \rho_{2u}^2 - \rho_{12}^2 + 2\rho_{1u}\rho_{2u}\rho_{12};$$

and $\rho_{12} \leq \rho_{12}^u \leq 1$, $\rho_{12} \geq \rho_{12}^l \geq -1$. The integral on the left-hand-side of (C.9) is bounded since $-4 \leq 4\rho_{1u}\rho_{2u} \leq 4$.

With both terms in (C.3) being bounded, the posterior for $\rho_{12}|\rho_{1u}, \rho_{2u}$ is integrable.

References

- [1] Dawid, A.P., 1979. Conditional independence in statistical theory (with discussion). *The Journal of the Royal Statistical Society, Series B* 41, 1-31.
- [2] Ghosh, M, Ghosh, A., Chen, M.H., A. Angresti, 1999. Noninformative priors for one parameter item response models. Working paper, Department of Statistics, University of Florida, Gainesville.
- [3] Gradshteyn, I.S., I.M. Ryzhik, 2000. *Tables of Integrals, Series, and Products* (Academic Press, London).
- [4] Heckman, J., Honore, B., 1990. The empirical content of the Roy Model. *Econometrica* 58, 1121-1149.
- [5] Koop, G., D. J. Poirier, 1997. Learning about the across-regime correlation in switching regression models. *Journal of Econometrics* 78, 217-227.
- [6] Maddala, G.S., 1983. *Limited Dependent and Qualitative Variables in Econometrics* (Cambridge University Press).
- [7] Meng, X.L., D. van Dyk, 1997. The EM algorithm - and old folk-song sung to a fast new tune. *Journal of the Royal Statistical Society, Series B* 59, 511-567.
- [8] Poirier, D.J., 1995. *Intermediate Statistics and Econometrics: A Comparative Approach* (MIT, Massachusetts).
- [9] Poirier, D.J., 1998. Revising beliefs in nonidentified models, 1998. *Econometric Theory* 14, 483-509.
- [10] Poirier, D.J., P.A. Rudd, 1981. On the appropriateness of endogenous switching. *Journal of Econometrics* 16, 249-256.
- [11] Roy, A.D., 1951. Some thoughts on the distribution of earnings. *Oxford Economic Papers* 3, 135-146.
- [12] Sareen, S., 2000. *Estimation and Testing of Structural Parametric Sealed-Bid Auctions*. Ph.D thesis, University of Toronto.
- [13] Vijverberg, W.P.M., 1993. Measuring the unidentified parameter of the extended Roy model of selectivity. *Journal of Econometrics* 57, 69-89.
- [14] Woodbury, M.A., 1971. Discussion of paper by Hartley and Hocking. *Biometrics* 27, 808-817.

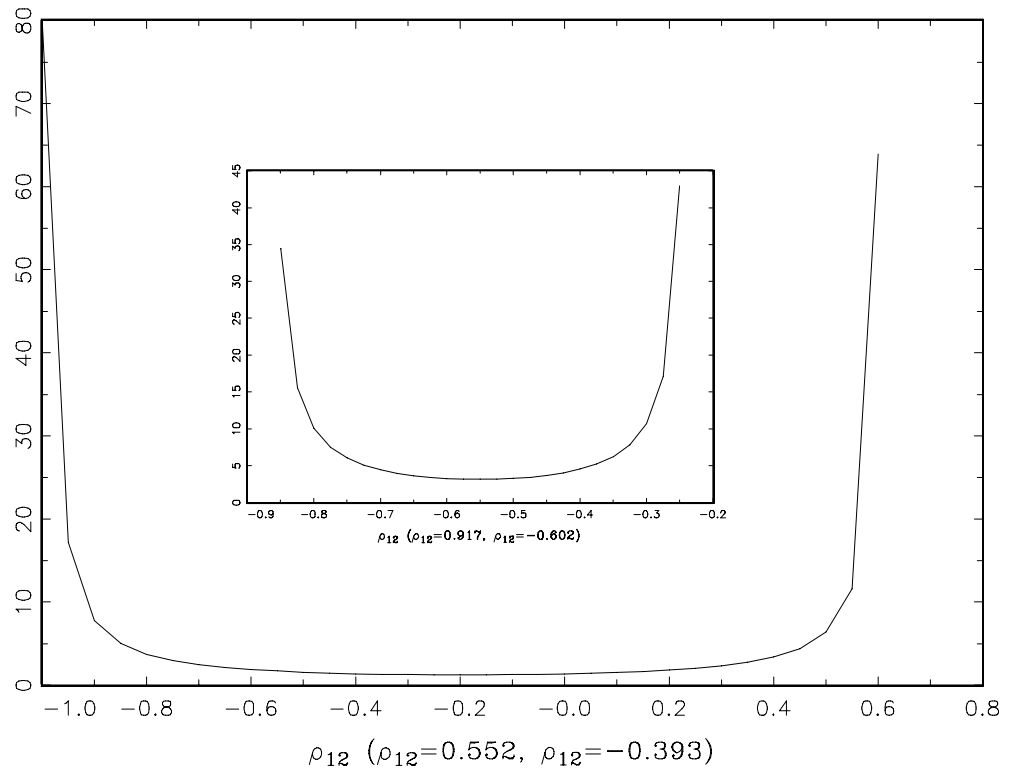


Figure 1: Posterior Kernel for $\rho_{12}|\rho_{1u}, \rho_{2u}$