Evaluating Data In Structural Parametric Auction, Job-Search And Roy Models*

Samita Sareen
Research and Risk Management
Bank of Canada, 234 Wellington
Ottawa, Ontario K1A 0G9
ssareen@bank-banque-canada.ca

June 25 2000

Abstract

Exact reference priors and posteriors are obtained for a structural, low-price, sealed-bid auction, job-search and Roy selection model. In all three models data is missing *but* with a difference; missing data is *unavailable* in job-search and Roy model unlike the auction model. The tails of the reference posterior based on the observed data are *flatter* compared to the reference posterior that is a product of the likelihood of observed data and the reference prior based on *both* the observed and the missing data. This follows from the observed data containing *strictly* less Fisher information about the parameters than both the observed and the missing data. For the Roy model, conditional inference about the unidentified coefficient of correlation of potential earnings in the two sectors is *possible* only if the reference prior is based on both the observed and missing data. As a result, when objective inference is being done in missing data problems, this paper suggests the use of reference priors based on both the observed and missing data.

JEL Classification: C11

Keywords: Job-search; Low-price, sealed-bid; Missing data; Reference prior; Roy Model

*I have benefitted from invaluable discussions with Subhashis Ghosal, Tony Lancaster, Mike Peters, Dale J. Poirier and participants at the International Workshop on Objective Bayesian methodology, Valéncia, 1999 and Sixth World meeting of the ISBA, Greece, 2000. All errors in interpretation are my responsibility.

1 Introduction

Missing data is encountered in several popular applications in economics. For example, in auction models either data on all bids within and across auctions or just winning bids across auctions are available. Paarsch (1992) studies first-price, sealed-bid tree planting auctions using data on winning bids. Laffont, Ossard and Vuong (1995) use winning bids to study a Dutch auction for eggplants. Bajari (1997) and Porter and Zona (1983) use data on all bids to study procurement auctions. In the Tobit model the censored data is missing. In job search models (Lancaster, 1997) the accepted wage offers and the duration of unemployment of individuals are observed; their rejected wage offers and the number of rejected offers are missing. In the Roy model of selection (Roy, 1951) an individual's occupational choice and her realized earnings in this occupation of choice, are observed; her potential earnings in the sector that she did not choose are *not* observed. Defining "occupations" broadly, Heckman and Honore (1990) list several applications of the Roy model - a women's choice between market and nonmarket work, a worker's choice between union and nonunion employment, etc. Rubin (1974, 1978) and Angrist, Imbens and Rubin (1996) have promoted the Roy model in the program evaluation literature. Recently Sareen (1999b, Chapter 4) has been studying timber auctions in the county of Simcoe in Southern Ontario. The bidders who have sawmills located outside the county of Simcoe have the choice of bidding in Simcoe or in the counties surrounding Simcoe. Since all she observes is where they actually bid, the Roy model is being used to ascertain how different their bids would have been in the surrounding counties where they *did not* bid.

In some applications data is missing, even though it is available, due to the *inability* of the empirical researcher to obtain this missing data. Thus, in a first-price, sealed-bid auction to procure crude-oil that I studied previously (Sareen, 1999a), only the winning bid was made public; I was unable to obtain the losing bids even though these bids were observed by the authority that conducted these auctions. On the other hand in "missing data" problems as pioneered by Tanner and Wong (1987), the missing data is *unavailable*. Job search models, Tobit model and the Roy model fall in this category; there are many examples of "missing data" problems besides these three models. Within auction models, in a Dutch auction, where an auctioneer calls out an initial high and continuously lowers this price till one bidder accepts the current price, only the winning bid is observed. It is the format of the auction that is responsible for the observation of the winning bids in Dutch auctions.

In this scenario where missing data is encountered, a concern of an empirical researcher is whether she is *always* losing expected information about the parameters that characterize the distribution of the data when she is working with *observed* data only. The answer is yes unless the observed data are *sufficient* for the parameters. In the first part of the paper I quantify this loss in expected information in terms of the Fisher information. I also illustrate this through structural auction, job-search and Roy models of selection; the structural auction model that I use for illustration is a parametric low-price, sealed-bid auction under the independent-private-values paradigm using parametric assumptions made by Paarsch (1992) for the distribution of private values.

A subjective Bayesian carrying out inference about the parameters of the distribution of private signals of the bidders will specify the same priors for the parameters irrespective of the data that is available. Inference about the parameters, since it is based on the posterior distribution of the parameters, will be affected through the likelihood function. If the observed data are not sufficient for the parameters, the likelihood function, and as a result, the posterior distribution of the parameters will be different depending on the data that is available.

In many applications specifying subjective priors is a difficult exercise. An example is the low-price, sealed-bid auction under the independent private-values paradigm and job search models. One reason is that the support of the data depends on the parameters of the distribution of the private signals of the bidders in structural auction models.[1] Specifying subjective priors for the parameters of the distribution of private signals of the bidders *implies* a prior on the support of the data since the support of the data is a function of these parameters. This implied prior has to be taken into consideration in any prior elicitation about the parameters of the distribution of the private signals of the bidders.[2]

Even if subjective priors are not difficult to specify, a subjective Bayesian may want to find structural rules that determine priors that are noninformative as a reference point in a prior sensitivity analysis.

Noninformative priors are appealing to both nonsubjective Bayesians and frequentists as well because the posterior probabilities based on certain noninformative priors agree with sampling probabilities to an order (Ghosh (1994) and Ghosh and Mukherjee (1992)).

In a pioneering paper, Bernardo (1979) has introduced the so-called reference prior to represent the idea of a noninformative prior. The reference prior emerges from maximizing an asymptotic expansion of Lindley's measure of information. Lindley's (1965) measure of information is defined as the expected Kulback-Liebler divergence between the posterior and the prior; the larger the measure, more informative the data and hence less informative the prior. When there are no nuisance parameters and

[1] For example, in a Dutch auction under the independent-private-values paradigm the bids or the winning bid are less than the unconditional expected value of the second highest private value. In a low-price, sealed-bid auction under the independent-private-values paradigm, the bids or the winning bid are greater than the equilibrium bid that corresponds to a zero cost of procuring the object being auctioned. In job search model the accepted wage offer is greater than, and the rejected offers of an individual less than the reserve wage.

[2] This was one reason why conjugate priors were used in Sareen (1999). Since priors are conjugate with respect to a likelihood function, the nonregular feature of the likelihood function in that the support of the data depends on the parameters of the distribution of the private signal of the bidders, is taken account of in prior elicitation of these parameters.

certain regularity conditions are satisfied, Bernardo's reference prior is the Jeffreys' prior. The Roy model of selection is an example. Ghosh (1994), Ghosal and Samanta (1995) and Ghosal (1997), extending Bernardo's (1979) work, have obtained the reference prior when the support of the data depends on the parameters; job-search and structural parametric auction models are examples in this category. In general, in this scenario, the reference prior is *not* the Jeffreys' prior.

Reference priors depend on the sampling density of the data. In the scenario when data is missing an important feature of the sampling density of the data is the loss in expected information about the parameters that characterize the data distribution unless the observed data are sufficient for the parameters. In the second part of the paper I *link* this loss to the specification of reference priors for structural parametric auction, job search, and Roy selection model.

When the observed data are sufficient for the parameters that characterize the data distribution, the reference prior as well as the *exact* reference posterior distributions, whether obtained from the observed data or the observed and the missing data, are identical. Hence inference about the parameters is unaffected by whether or not, the missing data is available. This scenario is a special case.

In most applications the observed data are *only* members of the set of sufficient statistics. As a result, the expected information about the parameters from the observed data is *strictly less* compared to when the missing data is observed as well. In this scenario, I show that the reference prior and the *exact* reference posterior distribution of the parameters obtained from the observed data is more dispersed than that obtained from the observed and the missing data. The message to an empirical researcher is to obtain the data that is missing and base inference on *all* the data. It *may* be feasible to obtain the data that is not observed in sealed-bid auctions since both the winner and the losing bidders submit bids to the seller. In "missing data" problems like the job-search and the Roy models of selection it is *not* feasible to obtain the missing data since it is *unavailable*. In this scenario using the reference prior based on *both* the observed and missing data may be attractive. I show through examples that the tails of the the exact reference posterior based on this prior provides more information about the parameter than the reference posterior that is a product of the likelihood of the observed data and the reference prior obtained from *only* the observed data. In addition, I show that inference about the unidentified coefficient of correlation between earnings in the sectors of choice, conditional on the identified parameters, is possible *only* if the reference posterior is the product of the likelihood of the observed data and the reference prior obtained from both the missing and observed data.

An outline of the paper is as follows. Section 2 describes the statistical framework used in the paper. I also describe the three models that will be used to illustrate the basic results of this paper. The first example is a low-price, sealed-bid auction under the specification of the distribution of private values as in Paarsch (1992). The job-search model as described in Lancaster (1997) is the second example. The third example is the Roy model of selection; the discussion of the Roy model closely follows Koop and Poirier (1997). Section 3 proves that when the observed data are not sufficient for the parameter vector, the difference in the Fisher information about the parameters obtained from the observed and the missing data compared with the observed data only is a *non-null psd* matrix. I illustrate this result with the three models described in Section 2. A brief review of the concept of reference prior is given in Section 4. In Sections 5, 6 and 7, I discuss reference analysis for a low-price, sealed-bid auction, job-search, and the Roy selection model, respectively. Section 8 concludes.

2 Statistical Framework

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. The quantities used in this paper are defined in Table 1; all quantities defined in Table 1 are evaluated at a value θ_o of θ; I will indicate when θ_o stands for the "true" value of θ. Given a random variable $\mathbf{V}\epsilon\Re^+$, $E_{\mathbf{V}|\theta}(\bullet)$ and $Var_{\mathbf{V}|\theta}(\bullet)$ will indicate the expected value and the variance, respectively, with respect to the distribution $\mathbf{V} \mid \theta$. The support of the density function of a random variable $\mathbf{V}$ is denoted by $\zeta_{\mathbf{V}}$; if the support depends on θ it will be indicated by $\zeta_{\mathbf{V}}(\theta)$. $\pi(\theta)$ and $\pi(\theta|\mathbf{v})$ will indicate the prior and the posterior density functions of θ, respectively.

Assumption 1: I will indicate the q dimensional data vector by $\mathbf{V}$. The data could be either univariate or multivariate. In case it is multivariate, $\mathbf{V}$ will refer to the vector formed by "stacking" the observations on each variable. The subscript j will be used to represent an element of $\mathbf{V}$; thus, V_j is the jth element of $\mathbf{V}$. The distribution of $\mathbf{V}$ is characterized by a $k \times 1$ parameter vector $\theta \in \mathbf{\Theta}$, with density function $f_{\mathbf{v}}(\mathbf{V} \mid \theta)$. $f_v(V_j \mid \theta)$ is the density function of the jth element of $\mathbf{V}$. I will also be referring to an arbitrary partitioning of the q dimensional data vector $\mathbf{V}$ into a p dimensional vector of statistics, $\mathbf{V}_p^O$, and a $q-p$ dimensional vector of statistics, $\mathbf{V}_{q-p}^M$, with $p < q$ and $p \geq k$. Thus $\mathbf{V} = [\mathbf{V}_p^O, \mathbf{V}_{q-p}^M]'$. In the three models that I use as examples, and that are described below, I will indicate the ***observed*** data by $\mathbf{V}_p^O$ and the ***missing*** data by $\mathbf{V}_{q-p}^M$.

When standard regularity conditions (see Poirier, 1995, p. 259) are satisfied so that the support of the data does not depend on θ, I will refer to it as the ***regular*** case. If the ***support of the data depends on the parameter*** θ, but other regularity conditions are satisfied, I will refer to it as the ***nonregular*** case. For the nonregular case I make the following assumption if θ is a vector.

Assumption 2: Assuming θ is a vector and can always be partitioned into $\theta = (\eta, \varphi)$, where η is a scalar and φ is a $k-1$ dimensional vector,
(a) the support of the density function of the random variable V_j, $f_v(v_j|\theta)$, depends on θ, $V_j \geq l(\eta, \varphi)$;

(b) $l(\eta, \varphi)$ is a strictly monotonic function of η for every φ.

I will refer to the ***regular parameter by*** φ and the ***nonregular parameter by*** η. The importance of Assumption 2 is that it enables me to rewrite $f_v(v_j|\eta, \varphi)$ as $f_v(v_j|l^{-1}(V_*), \varphi)$. Then evaluating all quantities at the "true" value of φ, the expected value of the score function of the log likelihood of bids with respect to φ is zero, $E_{\mathbf{V}|\theta}[\mathbf{S_V}(\varphi_o; \mathbf{V})] = 0$; and the Fisher information for φ equals the negative of the sampling expectation of the Hessian matrix , $E_{\mathbf{V}|\theta}[\mathbf{S_V}(\varphi_o; \mathbf{V})\mathbf{S}'_{\mathbf{V}}(\varphi_o; \mathbf{V})] = -E_{\mathbf{V}|\theta}[\mathbf{H_V}(\varphi_o; \mathbf{V})]$.[3] These results hold for the entire parameter vector θ when standard regularity conditions hold since regularity allows the interchange of the operations of differentiation and integration. They break down when the support of the data depends on θ. By fixing the lower boundary of the distribution of bids, the dependence of the support of the density function of the bids on θ is removed, making the likelihood function "regular" with respect to φ again.

I next specify the three models that will be used to illustrate the results in this paper. The first two, structural auction and job-search models are "nonregular" models since the support of the data depends on the parameters. The Roy selection model is a "regular" model.

2.1 Low-Price, Sealed-Bid Auction

Low-price, sealed-bid auctions under the symmetric independent-private-values paradigm have been previously studied by Paarsch (1992) and Sareen (1999). While Paarsch studies tree planting auctions in British Columbia, Sareen studies auctions for procurement of crude-oil on the international market by the Indian Oil Corporation, a public sector undertaking in India. I will indicate an auction by the subscript j, with $j = 1, ...T$; the superscript i will indicate a bidder in an auction, with $i = 1, ..., n$. C_j^i is the private cost of bidder i in auction j; it is assumed to follow some parametric distribution characterized by the parameter vector $\theta = (\delta, \varphi)$. B_j^i is the bid of bidder i in auction j. The nT dimensional column vector, $\mathbf{B} = [\mathbf{B}_1, .., \mathbf{B}_j, ..., \mathbf{B}_T]'$, is the vector of bids for the n bidders in the T auctions; $\mathbf{B}_j$ is an n dimensional column vector of n bids in auction j, $\mathbf{B}_j = [B_j^1, .., B_j^i, ..., B_j^n]'$. In this model $\mathbf{B} = \mathbf{V}$ comprises the observed and the missing data with $q = nT$. I also assume that $B_j^i|\theta$ is independently and identically distributed across the n bidders and the T auctions.

W_j is the winning bid in auction j; this is the minimum winning bid in auction j. W_* is the minimum winning bid across auctions $j = 1, ..., T$, $W_* = \min_{j=1,...,T} W_j$. The T dimensional vector of winning bids, $\mathbf{W} = [W_1, ..., W_T]'$ comprises the observed data; $\mathbf{W} = \mathbf{V}_p^O$, with $p = T$. The $n - 1$ ordered bids, *excluding* the winning bid, in each of the j auctions comprises the $q - p$ dimensional vector of missing data. Thus $\mathbf{V}_{q-p}^M = \mathbf{B}_{n-1:n}$,where $\mathbf{B}_{n-1:n} = [\mathbf{B}_{1,n-1:n}, ..., \mathbf{B}_{T,n-1:n}]$; $\mathbf{B}_{j,n-1:n} = [B_j^{1:n}, ..., B_j^{n-1:n}]'$ is the $n-1$ dimensional vector of ordered bids in auction j excluding the winning bid.

[3] For a proof see Hong (1998).

I have reparametrized the lower bound of the support of the density function of bids, winning bids and the minimum winning bid, $l(\delta, \varphi, n)$, as $\eta(\delta, \varphi, n)$, where n is the number of potential bidders. I am assuming n is the same as the number of participants in an auction.[4] φ is the "regular" parameter and η or δ the "nonregular" parameter.

When bids are assumed to be proportional to costs, then Paarsch (1992, p. 203) assumes that C_j^i follows a Pareto distribution; $C_j^i \sim Pa(\delta, \varphi)$ with $C_j^i > \delta$, where Pa indicates a Pareto distribution.[5] Let $\eta(\delta, \varphi, n) = \frac{\delta\varphi(n-1)}{\varphi(n-1)-1}$ indicate a function of δ and φ; the dependence of $\eta(\delta, \varphi, n)$ on δ and φ will be suppressed for notational convenience. The equilibrium bid follows a Pareto distribution, $B_j^i \sim Pa(\eta, \varphi)$, $B_j^i > \eta$, with density function

$$f_b(b_j^i|\eta, \varphi) = \frac{\varphi\eta^{\varphi}}{(b_j^i)^{\varphi+1}}, \qquad b_j^i > \eta. \tag{1}$$

The winning bid follows a Pareto distribution too; $W_j \sim Pa(\eta, n\varphi)$, $W_j > \eta$ with density function

$$f_w(w_j|\eta, n\varphi) = \frac{n\varphi\eta^{n\varphi}}{(w_j)^{n\varphi+1}}, \qquad w_j > \eta. \tag{2}$$

The minimum winning bid across the T auctions, W_*, follows a Pareto distribution, $W_* \sim Pa(\eta, nT\varphi)$ with density function

$$f_{w_*}(w_*|\eta, nT\varphi) = \frac{nT\varphi\eta^{n\varphi}}{(w_*)^{nT\varphi+1}}, \qquad w_* > \eta. \tag{3}$$

φ is the shape parameter of the distributions of C_j^i, B_j^i, W_j, and W_*. δ is some minimum cost; it has an interpretation similar to δ in the previous case for the low-price, sealed-bid auctions. Thus δ is some minimum common cost of procuring crude oil that is known to all the bidders. $\frac{\varphi(n-1)}{\varphi(n-1)-1}$ is the per unit increase in a bidder's equilibrium bid as the cost of procuring oil increases.[6] $\frac{\varphi(n-1)}{\varphi(n-1)-1}\delta$ or η is the winning bid when the bidder's cost of procuring crude oil is the minimum cost known to all the bidders, δ.

---

[4]Justification for this assumption is given in Sareen (1999b, pp. 57-58, footnote 6).

[5]δ, η is measured in U.S. dollar per metric tonne since C_j^i is measured in this unit. φ is a unitless quantity; Johnson, Kotz and Balakrishnan (1994, p. 573) refer to it as *Pareto's constant* or a shape parameter.

[6]For a low-price, sealed-bid symmetric independent-private-values auction, Paarsch (1992, p. 195) has shown that

$$b_j^i = c_j^i + \frac{\int_{c_j^i}^{\infty} [F_c(\xi)]^{n-1} d\xi}{[F_c(c_j^i)]^{n-1}},$$

where $F_c(\bullet)$ is the distribution function of C_j^i. Since $C_j^i \sim Pa(\delta, \varphi)$, the distribution function of C_j^i is $F_c(c_j^i \mid \delta, \varphi) = 1 - \left(\frac{\delta}{c_j^i}\right)^{\varphi}$. Hence $b_j^i = \frac{\varphi(n-1)}{\varphi(n-1)-1}c_j^i$.

2.2 Job Search Model

I will follow the exposition of stationary job search models as in Lancaster (1997) with optimality *not* enforced.[7] An agent is indicate by the subscript j; there are $j = 1, ..., N$ agents. Under *full observability*, the following are observed for each agent j: (a) the number of rejected offers, s_j; (b) the rejected wage offers, $\mathbf{u}_j = [u_j^1, ..., u_j^{s_j}]'$; (c) the accepted wage, w_j; (d) the duration of unemployment, t_j. The q dimensional vector of observed and missing data is $\mathbf{V} = [\mathbf{S}, \mathbf{U}, \mathbf{W}, \mathbf{T}]'$ with $q = \sum_{j=1}^{N} s_j + 3N$. $\mathbf{S} = [S_1, ..., S_N]'$ is the vector of rejected offers of the N agents, and $\mathbf{U} = [\mathbf{U}_1, ..., \mathbf{U}_N]'$ are these rejected wage offers. $\mathbf{W} = [W_1, ..., W_N]'$ is the vector of accepted wage offers of the N individuals; $\mathbf{T} = [T_1, ..., T_N]'$ is the duration of unemployment of the N agents. From Lancaster (1997, p. 167) the likelihood function under *full observability* is the product over the $j = 1, ..., N$ individuals of the density function,

$$f_{suwt}(s_j, \mathbf{u}_j, w_j, t_j | \eta, \varphi) = \lambda^{s_j+1} e^{-\lambda t_j} fw(w_j|\mu, \sigma) \prod_{i=1}^{s_j} fw(u_j^i|\mu, \sigma)/s_j!,$$

$$w_j \geq \xi; \qquad u_j^1, ..., u_j^{s_j} < \xi. \tag{4}$$

$fw(\bullet|\mu, \sigma)$ is the density function of wage offers; Lancaster assumes that $\log W_j$ follows a Normal distribution with parameters μ, σ. ξ is the reserve wage; thus $\eta = \xi$ is the "nonregular" parameter. $\varphi = [\lambda, \mu, \sigma, \rho]'$ are the "regular" parameters.

Under *partial observability* only the accepted wage offer and the duration of unemployment are observed; that is, $\mathbf{V}_p^O = [\mathbf{W}, \mathbf{T}]'$ is the *observed* data, with $p = 2N$. The *missing* data is the number of rejected offers and these rejected wage offers for each individual; thus $\mathbf{V}_{q-p}^M = [\mathbf{S}, \mathbf{U}]'$. The likelihood function under partial observability is the product over the $j = 1, ..., N$ agents of the density function (Lancaster, 1997, p. 167)

$$f_{wt}(w_j, t_j | \eta, \varphi) = \lambda e^{-\lambda \overline{F}_w(\xi|\mu,\sigma) t_j} fw(w_j|\mu, \sigma), w_j \geq \xi. \tag{5}$$

I will indicate the minimum accepted wage across the N individuals by W_*; that is $W_* = \min_{j=1,...,N} W_j$. U_* is the largest rejected wage offer across the N individuals; $U_* = \max_{i=1,...,s_{j_{j=1,...,N}}} U_j^i$.

2.3 Roy Selection Model

The description of the Roy model closely follows Koop and Poirier (1997). $j = 1, ..., n$ individuals face a choice between two "sectors".[8] The potential earnings of each

[7] If optimality is enforced the reserve wage, which is the "nonregular" parameter, is a determinstic function of the other parameters through the optimality condition (Lancaster, p. 167, 1997). Since the focus is on the "nonregular" parameter in this paper, this is not an interesting case.

[8] The 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. In the timber auctions that I am studying currently the two sectors will refer to participating in the county of Simcoe or in the counties neighbouring Simcoe.

individual i in sectors 1 and 2 is given by Y^*_{1j} and Y^*_{2j}, respectively. $\mathbf{X}_{1j}$ and $\mathbf{X}_{2j}$ are the covariates affecting potential earnings in sectors 1 and 2, respectively, that are observed by the econometrician; these are of dimension $k_1 \times 1$ and $k_2 \times 1$, respectively. These could be personal characteristics measuring skill-levels in each sector, as well as, sector specific characteristics. The factors affecting potential earnings in the sector 1 and 2 that are unobserved by the econometrician 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. I^*_j indicates the difference in utility experienced by individual j when she works in sector 1 versus sector 2. This is affected by covariates $\mathbf{Z}_j$ observed by the econometrician and factors u_j unobserved by the econometrician. If $I^*_j \geq 0$, individual j will choose sector 1; if $I^*_j < 0$, individual j will choose sector 2.[9]

Omitting subscripts for individuals, the Roy model is

$$I^* = \mathbf{W}\gamma + u, \qquad Y^*_1 = \mathbf{X}_1\beta_1 + \epsilon_1, \qquad Y^*_2 = \mathbf{X}_2\beta_2 + \epsilon_2. \tag{6}$$

In the terminology of this paper $\mathbf{V}$, the vector of both the observed and missing data is $\mathbf{V} = [\mathbf{I}^*, \mathbf{Y}^*_1, \mathbf{Y}^*_2]$; $\mathbf{I}^*, \mathbf{Y}^*_1$ and $\mathbf{Y}^*_2$ are each n dimensional vectors. I will refer to this scenario as *full observability*. Thus under *full observability*, the following are observed for each of the $j = 1, ..., n$ agents: (a) potential earnings in sectors 1 and 2, Y^*_{1j}, Y^*_{2j}, respectively; and (b) the difference in utility between sector 1 and 2, I^*_j.

Following convention I will assume $j = 1, ..., n$ independently and identically distributed individuals;[10] for each individual j,

$$(I^*_j, Y^*_{1j}, Y^*_{2j})' \sim N_3(\mu_j, \mathbf{\Omega}), \tag{7}$$

where $\mu_j = [\mathbf{W}_j\gamma,\ \mathbf{X}_{1j}\beta_1,\ \mathbf{X}_{2j}\beta_2]'$ and

$$\mathbf{\Omega} = \begin{bmatrix} 1 & \rho_{1u}\sigma_1 & \rho_{2u}\sigma_2 \\ \rho_{1u}\sigma_1 & \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{2u}\sigma_1 & \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}. \tag{8}$$

The likelihood function under *full observability* is the product over $j = 1, ..., n$ individuals of the density function of this trivariate Normal distribution;

$$f_{I^*, Y^*_1, Y^*_2}(\mathbf{z}_j|\theta) = (2\pi)^{-1/6}|\mathbf{\Omega}|^{-1}e^{-\frac{1}{2}\mathbf{z}'_j\mathbf{\Omega}^{-1}\mathbf{z}_j}. \tag{9}$$

$\theta = \varphi = [\gamma, \beta_1, \beta_2, \sigma_1, \sigma_2, \rho_{1u}, \rho_{2u}, \rho_{12}]'$ since the Roy model is a regular model; and $\mathbf{z}_j = [I^*_j - \mathbf{W}_j\gamma, Y^*_{1j} - \mathbf{X}_{1j}\beta_1, Y^*_{2j} - \mathbf{X}_{2j}\beta_2]'$. I will also refer to the parameter vector $\varphi = [\varphi_*, \rho_{1u}, \rho_{2u}, \rho_{12}]'$, where $\varphi_* = [\gamma, \beta_1, \beta_2, \sigma_1, \sigma_2]'$.

[9] Roy (1951) assumed that an individual's sectoral choice is based exclusively on sectoral earnings.

[10] For example see Roy (1953), Poirier and Rudd (1981), Madalla (1987), Heckman and Honore (1990), Vijverberg (1993) and Koop and Poirier (1997).

Of the n agents I 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

$$\mathbf{Y}_1^* = \begin{bmatrix} \mathbf{Y}_1^O \\ \mathbf{Y}_1^M \end{bmatrix}, \quad \mathbf{Y}_2^* = \begin{bmatrix} \mathbf{Y}_2^O \\ \mathbf{Y}_2^M \end{bmatrix}, \tag{10}$$

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, all that an econometrician *ever* observes is this individual's realized earnings in the sector of choice; these will be indicated by Y_{1j}^O or Y_{2j}^O depending on whether sector 1 or sector 2 is chosen by individual j. An econometrician will *never* observe the potential earnings of individual j in the sector that she did *not* choose. The econometrician observes the following for each of the j individuals,

$$Y_{1j}^O = \begin{cases} Y_{1j}^* & \text{iff} \quad I_j^* > 0 \\ 0 & \text{otherwise.} \end{cases} \quad Y_{2j}^O = \begin{cases} Y_{2j}^* & \text{iff} \quad I_j^* \leq 0 \\ 0 & \text{otherwise.} \end{cases} \quad I_j = \begin{cases} 1 & \text{iff} \quad I_j^* > 0 \\ 0 & \text{iff} \quad I_j^* \leq 0. \end{cases} \tag{11}$$

This scenario will be refereed to as *partial observability*. Under *partial observability* the following are observed for each agent $j = 1, ..., n$: (a) realized earnings in the sector of choice, Y_{1j}^O or Y_{2j}^O; (b) sectoral choice, I_j. In the terminology of this paper, the vector of observed data $\mathbf{V}_p^O = [\mathbf{Y}_1^O, \mathbf{Y}_2^O, \mathbf{I}]'$, where $\mathbf{I} = [\mathbf{1}_{n_1}, \mathbf{0}_{n_2}]'$ is an n dimensional vector. The vector of missing data is $\mathbf{V}_{q-p}^M = [\mathbf{Y}_1^M, \mathbf{Y}_2^M, \mathbf{I}^*]'$. The likelihood function under *partial observability* is

$$L(\theta; \mathbf{Y}_1^O, \mathbf{Y}_2^O, \mathbf{I}) = \prod_{j=1}^{n} \left[\int_0^\infty f_{1u}(y_{1j}^O, u_j|\theta)\right]^{I_j} \left[\int_{-\infty}^0 f_{2u}(y_{2j}^O, u_j|\theta)\right]^{1-I_j}, \tag{12}$$

where $f_{1u}(\bullet|\theta)$ and $f_{2u}(\bullet|\theta)$ are density functions of bivariate normal distributions that can be obtained from the trivariate normal distribution specified in (7) (Poirier, 1995, p. 122). Note that ρ_{12} does not enter the likelihood function under partial observability.

3 Relating Fisher Information

In this section I first establish the relationship between the Fisher information about θ from both the observed and missing data with the Fisher information about θ from *only* the observed data. The Fisher information from the observed data is *identical* to the Fisher information from the observed and missing data *if and only if* the observed data is sufficient for θ.

These results are next illustrated with the structural auction, job-search and Roy models described in the last section.

For the structural low-price, sealed-bid auction described in the last section, given the "regular" parameter, the minimum winning bid is sufficient for the *scalar* "nonregular" parameter. For the "regular" parameters, the winning bids (or the minimum winning bid) are just members of the set of sufficient statistics, so that the expected information about the parameters is *strictly less* if just the winning bids are used.[11]

Job-search model is *like* the structural auction model in that the support of the data depends on the parameters. Specifically, the accepted wage offer of an individual is greater than the reserve wage, the "nonregular" parameter, and all the rejected wage offers of an individual are less than the reserve wage. *Unlike* the "nonregular" parameter in structural auction models, the Fisher information about the reserve wage from the observed data on the accepted wage offers and the duration of unemployment is *strictly* less than the Fisher information about the reserve wage when data on the rejected wage offers and the number of these rejected offers is available, as well.

In the Roy model the coefficient of correlation between the potential earnings in the two sectors is not identified under *partial observability.* The data observed under *partial observability* comprises sectoral choice and realized earnings in the sector of choice. This has been documented by Heckman and Honore (1990), Vijverberg (1993) and Koop and Poirier (1997). Vijverberg and Koop and Poirier also note that despite this lack of identification, data on realized earnings in the sector of choice contain expected information about the unidentified correlation coefficient. Expected information about the identified parameters from the data on realized earnings in the sector of choice "spills-over" to the nonidentified correlation coefficient through the positive definiteness restriction on the covariance matrix. This expected information is quantified in Example 3 below. I also show that this expected information about the correlation coefficient is *strictly less* than the expected information obtained under *full observability* when both the realized earnings in the sector of choice and the potential earnings in the sector not chosen are observed.

I first express the Fisher information from the q dimensional data vector $\mathbf{V}$, which comprises both the observed and the missing data, as the sum of two terms in the Lemma below. The first term is the Fisher information from the p dimensional vector of statistics, $\mathbf{V}_p^O$, the vector of observed data. The second term is the average Fisher information from the $q-p$ dimensional vector of statistics, $\mathbf{V}_{q-p}^M$, the vector of missing

[11] A similar concern arises in the study of natural phenomena in the statistics literature. For example the distribution of the highest wave is important in the design of sea structures, the distribution of the largest flood is important in designing dams, etc. The concern over the loss of information in using the largest observation has lead to the generalized Pareto Distribution (Castillo and Hadi, 1997 and Pickands, 1975) which is used to model exceedences over a threshold, thereby considering several of the largest observations instead of just the largest one.

data; the average is over $\mathbf{V}_p^O|\theta$, the sampling distribution of the p dimensional vector of statistics, $\mathbf{V}_p^O$.

Lemma 1: Under Assumption 1, using quantities defined in Table 1, and either
(a) θ is a k dimensional vector and standard regularity conditions hold; or
(b) $\theta = \varphi$ is a $(k-1)$ dimensional vector and Assumption 2 holds,

$$\mathbf{J}_{\mathbf{V}}(\theta_{\mathbf{o}}) = \mathbf{J}_{\mathbf{V}_p^O}(\theta_{\mathbf{o}}) + E_{\mathbf{V}_p^O|\theta}\left[\mathbf{J}_{\mathbf{V}_{q-p}^M|\mathbf{V}_p^O}(\theta_{\mathbf{o}})\right]. \tag{13}$$

All quantities have been evaluated at $\theta = \theta_{\mathbf{o}}$. $\mathbf{J}_{\mathbf{V}}(\theta_o)$ is the Fisher information from the data matrix $\mathbf{V}$ which comprises both the observed and the missing data. $\mathbf{J}_{\mathbf{V}_{q-p}^M|\mathbf{V}_p^O}$ is the Fisher information of the loglikelihood function of the distribution of $\mathbf{V}_{q-p}^M$, the missing data, conditional on $\mathbf{V}_p^O$, the observed data. $\mathbf{J}_{\mathbf{V}_p^O}(\theta_{\mathbf{o}})$ is the Fisher information from the p dimensional vector of statistics,$\mathbf{V}_p^O$, which is the observed data.

: See Appendix A.

I now show that the difference in the Fisher information about θ from the observed and the missing data compared with the Fisher information from *only* the observed data is a *null* matrix *if and only if* the observed data is *sufficient* for θ.

Corollary 1: From Lemma 1,

$$\mathbf{J}_{\mathbf{V}}(\theta_{\mathbf{o}}) = \mathbf{J}_{\mathbf{V}_p^O}(\theta_{\mathbf{o}}) \quad \text{iff} \quad \mathbf{V}_p^O \quad \text{jointly sufficient for} \quad \theta,$$

where all quantities have been evaluated at $\theta = \theta_o$.

See Appendix B.

Note that I have proved Lemma 1 and Corollary 1 *without* assuming that V_j are independently and identically distributed. In the low-price, sealed-bid auction, the bids, B_j^i are independently and identically distributed across bidders and auctions. In job search models, $W_j, \mathbf{U}_j, S_j, T_j$ are independently and identically distributed across the $j = 1, ..., N$ individuals; however, $W_j, \mathbf{U}_j, S_j, T_j$ are not independent. In the Roy model of selection, $[I^*, Y_1^*, Y_2^*]$ are not independent. But $[I^*, Y_1^*, Y_2^*]$ are independently, though not identically, distributed across the n individuals.

I now demonstrate Lemma 1 and Corollary 1 with the models described in Section 2.

Example 1: Low-price, sealed-bid auction
Assuming φ is given, the Fisher information about the "nonregular" parameter, η, from the minimum winning bid across auctions, W_*, is identical to the Fisher information about η from all the bids, $\mathbf{B}$.[12] That is, the minimum winning bid is sufficient for η given φ.[13] I have obtained the Fisher information matrix for (η, φ) from all the

[12]Since $\eta = \frac{\varphi(n-1)}{\varphi(n-1)-1}\delta$, identical results hold for δ as well.

[13]Another way of establishing the sufficiency of the minimum winnning bid for η *given* φ is due to Huzurbazar (1976, pp. 158, 174-186). He states the necessary and sufficient conditions under which the sample minimum is sufficient for *a* scalar "nonregular" parameter. For η *given* φ, the density function of bids given by equation (1) satisfies these necessary and sufficient conditions.

bids, winning bids across auctions and the minimum winning bid across auctions in Appendix C. Given φ, the Fisher information about η is the element in the first row and column of these matrices,

$$J_{\mathbf{B}}^{\eta\eta}(\eta,\varphi) = J_{\mathbf{W}}^{\eta\eta}(\eta,\varphi) = J_{W_*}^{\eta\eta}(\eta,\varphi) = \frac{(nT)^2\varphi^2}{\eta^2}. \tag{14}$$

The T dimensional vector of winning bids, $\mathbf{W}$, is however *not sufficient* for the "regular" parameter φ. From Appendix C, the Fisher information for φ from all the bids within and across auctions, and from the winning bids across auctions, is the element in the second row and column of the Fisher information matrices, $J_{\mathbf{B}}(\theta)$ and $J_{\mathbf{W}}(\theta)$, respectively,

$$\mathbf{J}_{\mathbf{B}}^{\varphi\varphi}(\eta,\varphi) = \frac{1}{\varphi^2}\left(nT + \frac{(nT)^2\varphi^2}{(\varphi m-1)^2}\right) \quad \text{and} \tag{15}$$

$$\mathbf{J}_{\mathbf{W}}^{\varphi\varphi}(\theta,\varphi) = \frac{1}{\varphi^2}\left(T + \frac{(nT)^2\varphi^2}{(\varphi m-1)^2}\right). \tag{16}$$

Clearly $\mathbf{J}_{\mathbf{B}}^{\varphi\varphi}(\eta,\varphi) > \mathbf{J}_{\mathbf{W}}^{\varphi\varphi}(\theta,\varphi)$; hence the winning bids provide *strictly less* information about φ than the bids of all the n bidders across the T auctions.

Example 2: Job Search Models

In job search models I focus on the "nonregular" parameter, the reserve wage ξ; that is, $\eta = \xi$. I focus on ξ because *like* the structural auction model in Example 1, ξ is the "nonregular" parameter in the job-search model. However, *unlike* the structural auction model is *not sufficient* for the reserve wage ξ; the sample minimum, W_*, is the minimum accepted wage across the N agents. The intuition for this result is simple. Since the reserve wage ξ is greater than the largest rejected wage offer across the N agents, the missing data on the rejected wage offers provides expected information about the reserve wage too. Making use of Lemma 1, I obtain, in Appendix D, the difference in the Fisher information about ξ from the observed data, as well as, the missing data and the Fisher information from the observed data only, $E_{\mathbf{W},\mathbf{t}|\theta}\left[J_{\mathbf{U},\mathbf{S}|\mathbf{W},\mathbf{t}}^{\xi\xi}(\theta)\right]$,

$$E_{\mathbf{W},\mathbf{t}|\theta}\left[J_{\mathbf{U},\mathbf{S}|\mathbf{W},\mathbf{t}}^{\xi\xi}(\theta)\right] = N(N+1)\left[\lambda f_w(\xi|\mu,\sigma)\right]^2\left[\lambda\ \bar{F}_w\ (\xi|\mu,\sigma)\right]^2. \tag{17}$$

Since $E_{\mathbf{W},\mathbf{t}|\theta}\left[J_{\mathbf{U},\mathbf{S}|\mathbf{W},\mathbf{t}}^{\xi\xi}(\theta)\right] > 0$, the difference in the Fisher information from all the data compared with just the observed data, $J_{\mathbf{U},\mathbf{S},\mathbf{W},\mathbf{t}}^{\xi\xi}(\theta) - J_{\mathbf{W},\mathbf{t}}^{\xi\xi}(\theta) > 0$.

Example 3: Roy Model of Selection

I will focus on the coefficient of correlation between potential earnings in the two sectors of choice, ρ_{12}. This is not identified from the observed data on on realized earnings and sectoral choice of individuals; ρ_{12} does not *explicitly* enter the likelihood function under *partial observability* given by equation (12). 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 the two sectors to be identified, the earnings of some individuals in *both* sectors have to be observed.

But as Vijverberg (1993, p. 71) points out, the positive definiteness of $\mathbf{\Omega}$ 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

$$\Psi(\rho_{1u}, \rho_{2u}) = \left[\rho_{12}^{l}, \rho_{12}^{u}\right] = \left\{\rho_{12}|\rho_{12} \in \left[\rho_{1u}\rho_{2u} \mp \sqrt{(1-\rho_{1u}^2)(1-\rho_{2u}^2)}\right]\right\}; \tag{18}$$

$\rho_{12}^{l}, \rho_{12}^{u}$ 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. Since ρ_{12} does not enter the likelihood function in (12), this expected information, $J^{\rho_{12}\rho_{12}}_{\mathbf{I},\mathbf{Y}_1^o,\mathbf{Y}_2^o}(\theta)$, cannot be obtained from the likelihood function. However Lemma 1 provides a way of quantifying this expected information,

$$J^{\rho_{12}\rho_{12}}_{\mathbf{I},\mathbf{Y}_1^o,\mathbf{Y}_2^o}(\theta) = J^{\rho_{12}\rho_{12}}_{\mathbf{I}^*,\mathbf{Y}_1^*,\mathbf{Y}_2^*}(\theta) - E_{\mathbf{I},\mathbf{Y}_1^o,\mathbf{Y}_2^o}\left[J^{\rho_{12}\rho_{12}}_{\mathbf{I}^*,\mathbf{Y}_1^M,\mathbf{Y}_2^M|\mathbf{I},\mathbf{Y}_1^o,\mathbf{Y}_2^o}(\theta)\right].$$

The first term is the Fisher information about ρ_{12} from data under *full observability*,

$$J^{\rho_{12}\rho_{12}}_{\mathbf{I}^*,\mathbf{Y}_1^*,\mathbf{Y}_2^*}(\theta) = \frac{1}{|\mathbf{\Omega}|}\left[1 + \frac{2}{|\mathbf{\Omega}|}\left(\rho_{1u}\rho_{2u} - \rho_{12}\right)^2\right], \tag{19}$$

where $|\mathbf{\Omega}| = 1 - \rho_{1u}^2 - \rho_{2u}^2 - \rho_{12}^2 + 2\rho_{1u}\rho_{2u}\rho_{12}$, is the determinant of $\mathbf{\Omega}$; this has been derived in Appendix F. The second term is the average Fisher information about ρ_{12} from the missing data; the average is over the sampling distribution of the observed data. The missing data comprises the potential earnings of an agent j in the sector she did not choose, Y_{1j}^M or Y_{1j}^M, and the difference in utility experienced by an agent from working in the two sectors, I_j^*. It is through the latter that ρ_{12} enters the likelihood function of the missing data.[14]

4 Review of Reference Priors

In this Section I provide an overview of the reference prior idea of Bernardo (1979). I begin by defining Lindley's measure of information which provides a measure of the information about the parameter in a sample and forms the basis for obtaining reference priors. The first two definitions are from Kass and Wasserman (1996, pp. 1345-1351).

Definition 1: **Lindley's measure of information** is the expected Kullback-Leibler distance between the posterior density, $\pi(\theta|\mathbf{z})$, and the prior density, $\pi(\theta)$,

$$I\left(\pi(\theta), \mathbf{Z}\right) = \int K\left(\pi(\theta|\mathbf{z}), \pi(\theta)\right) f(\mathbf{z}) d\mathbf{z},$$

[14]Koop and Poirier (1997, pp 222-223) have obtained the likelihood function of the missing data conditional on the observed data, $\mathbf{I}^*, \mathbf{Y}_1^M, \mathbf{Y}_2^M|\mathbf{I}, \mathbf{Y}_1^o, \mathbf{Y}_2^o$.

where $K(\pi(\theta|\mathbf{z}),\pi(\theta)) = \int_{\mathbf{\Theta}} \pi(\theta|\mathbf{z}) \log(\pi(\theta|\mathbf{z})/\pi(\theta))\, d\theta$, is the Kullback-Leibler distance between the posterior and the prior density. The expectation in Lindley's measure of information is with respect to the marginal density of the data, $f(\mathbf{z}) = \int_{\mathbf{\Theta}} f_{\mathbf{z}}(\mathbf{z}|\theta)\, \pi(\theta) d\theta$.

Given T independently and identically random variables, $\mathbf{Z}_T = (\mathbf{Z}_1,...,\mathbf{Z}_T)$, where $\mathbf{Z}_j$ is an n dimensional column vector, Lindley's measure of information given in Definition 1 is the asymptotic expected information that an experiment provides about θ. Hence as $T \to \infty$, $I_\infty^\pi = \lim_{T\to\infty} I(\pi(\theta), \mathbf{Z}_T)$ provides a measure of missing information as a function of the prior density function, $\pi(\theta)$ (Lindley, 1956 and Good, 1960, 1966); the larger the measure, the more informative the data and less informative the prior. Bernardo's idea was to maximize I_∞^π to obtain the reference prior. However I_∞^π involves the sample size and is, as a result, typically infinite. Bernardo got around this problem by first finding the sequence of priors $\pi_T(\theta)$ which maximize $I(\pi(\theta), \mathbf{Z}_T)$. Corresponding to this sequence of priors is a sequence of posterior density functions, $\pi_T(\theta|\mathbf{z}_T)$ with a limiting posterior $\pi(\theta|\mathbf{z})$; the reference prior is that positive function of θ that produces this limit posterior via Bayes' theorem.

The key idea behind the reference prior, whether in the regular or the nonregular case, is that Lindley's measure admits an asymptotic expansion with the leading term, free of $\pi(\theta)$, going to infinity with T, second term a function of $\pi(\theta)$ but free of T, and the remainder going to zero. Maximization of the second term, with respect to $\pi(\theta)$, gives the reference prior.

In the k dimensional regular case, Ibragimov and Has'minskii (1973) have obtained the asymptotic expansion of Lindley's measure of information,

$$\frac{k}{2} \log \frac{T}{2\pi e} + \int \pi(\theta) \frac{\sqrt{\det \mathbf{J}_Z(\theta)}}{\pi(\theta)} d\theta + o(1), \tag{20}$$

where $\mathbf{J}_Z(\theta)$ is the Fisher information from Z. Maximization with respect to $\pi(\theta)$ leads to Jeffreys' prior which is defined now.

Definition 2: **Jeffreys' prior**. Given data on some random variable Z, with density function $f_z(z|\theta)$, and $\mathbf{J}_Z(\theta)$ being the Fisher information matrix, Jeffreys' prior for θ, $\pi_{jeff}^z(\theta)$, is

$$\pi_{jeff}^z(\theta) \propto \sqrt{\det \mathbf{J}_Z(\theta)},$$

where "det" indicates the determinant of the matrix.

Jeffreys' prior is often used as a noninformative prior when the support of the data ***does not*** depend on θ. As $T \to \infty$ and the support of the data $\mathbf{Z}$ does not depend on θ, the posterior density for θ can be approximated by a Normal distribution with a variance-covariance matrix that equals the inverse of the Fisher information matrix evaluated at $\widehat{\theta}_{ml}$, the maximum likelihood estimator of θ, $\sqrt{T}\left(\theta - \widehat{\theta}_{ml}\right) \sim N\left(\mathbf{0}_k, \left[\mathbf{J}_Z\left(\widehat{\theta}_{ml}\right)\right]^{-1}\right)$. The sampling distribution of $\widehat{\theta}_{ml}$ as $T \to \infty$ is $\sqrt{T}\left(\widehat{\theta}_{ml} - \theta_o\right) \sim$

$N\left(\mathbf{0}_k, [\mathbf{J}_Z(\theta_0)]^{-1}\right)$, where θ_o is the "true" parameter value. Thus Jeffreys' prior is related to the variance-covariance matrix of the asymptotic distribution of the MLE or the asymptotic posterior distribution.

Corresponding to the asymptotic expansion of Lindley's measure of information in equation (21) for the regular case, Ghosal and Samanta (1997) have obtained an expansion of Lindley's measure of information for the one-parameter nonregular case,

$$\log \frac{T}{e} + \int \pi(\theta) \frac{|c(\theta)|}{\pi(\theta)} d\theta + o(1), \tag{21}$$

where

$$c(\theta) = E_{\mathbf{z}|\theta}\left[\frac{\partial}{\partial\theta} \log f_z(Z_j^i|\theta)\right] = l'(\theta) f_z\left(l(\theta)|\theta\right) - u'(\theta) f_z\left(u(\theta)|\theta\right). \tag{22}$$

$\zeta_Z(\theta) = [l(\theta), u(\theta)]$ defines the support of the sampling density of the data, $f_z(Z_j^i|\theta)$. Maximization of the second term in equation (22), with respect to $\pi(\theta)$, gives the reference prior for the scalar boundary parameter θ. The definition below is from Ghosal and Samanta (1997).

Definition 3: Reference prior for scalar parameter nonregular case.
(a) If θ is a scalar;
(b) $f_z(z|\theta)$, the sampling density of Z, is positive only on an interval $\zeta_Z(\theta)$ depending on θ, where $\zeta_Z(\theta) = [l(\theta), u(\theta)]$ and it is permitted that one of the end points need not depend on θ and may be plus or minus infinity; and
(c) $\zeta_Z(\theta)$ is monotone in θ;
the reference prior for θ, $\pi^z_{ref}(\theta)$, is

$$\pi^z_{ref}(\theta) \propto |c(\theta)|. \tag{23}$$

$|\quad|$ is the absolute value of $c(\theta)$ which I have defined in equation (23).

I now explain the role of $c(\theta)$. In a manner analogous to the regular case, as $T \to \infty$, the posterior distribution of θ is $T\left(\theta - \widehat{\theta}_{ml}\right) \sim EXP\left(c(\theta)\right)$ with variance $[c(\theta)]^{-2}$; $\widehat{\theta}_{ml} = \min\left[l^{-1}(W_*), u^{-1}(B^{n:n})\right]$, where W_* and $B^{n:n}$ are the minimum and maximum bid within and across auctions, respectively. The sampling distribution of $\widehat{\theta}_{ml}$ as $T \to \infty$ is $T\left(\widehat{\theta}_{ml} - \theta_o\right) \sim EXP\left(c(\theta_o)\right)$. Thus, like the regular case in Definition 2, the noninformative prior in the nonregular case is proportional to the square root of the precision of a suitably normalized function of the maximum likelihood estimator.

From Ghosal (1997), the reference prior when θ is a vector is defined next.

Definition 4: Reference prior for the multi-parameter non-regular case.

(a) Suppose θ is a vector, and $\theta = [\eta, \varphi]$, where η is a scalar, and φ a $k-1$ dimensional vector;
(b) further suppose $f_z(z|\theta)$ is positive only on an interval $\zeta_Z(\eta) = [l(\eta), u(\eta)]$ depending on η, with $\zeta_Z(\eta)$ being monotone in η and one of the end points need not depend on η.
Then the reference prior for $[\eta,\varphi]$, $\pi^z_{ref}(\eta, \varphi)$, is defined to be

$$\pi^z_{ref}(\eta, \varphi) \propto |c(\eta, \varphi)| \sqrt{\det \mathbf{J}_Z^{\varphi\varphi}(\eta, \varphi)}. \tag{24}$$

$c(\eta, \varphi)$ corresponds to the quantity given in equation (23) for the multi-parameter case,

$$c(\eta, \varphi) = E_{z|\eta,\varphi}\left[\frac{\partial}{\partial\eta} \log f_z(Z^i_j|\eta, \varphi)\right] = l'(\eta) f_z(l(\eta)|\eta, \varphi) - u'(\eta) f_z(u(\eta)|\eta, \varphi); \tag{25}$$

and $\mathbf{J}_Z^{\varphi\varphi}(\eta, \varphi)$ is the lower right hand block of $\mathbf{J}_Z(\eta, \varphi)$, the Fisher information matrix about η, φ from Z :

$$\mathbf{J}_Z^{\varphi\varphi}(\eta, \varphi) = E_{z|\eta,\varphi}\left[\frac{\partial}{\partial\varphi} \log f_z(Z^i_j|\eta, \varphi)\right]\left[\frac{\partial}{\partial\varphi} \log f_z(Z^i_j|\eta, \varphi)\right]'. \tag{26}$$

Since given η, the density function $f_z(z|\theta)$, is regular with respect to φ, I will refer to η as the "nonregular" parameter and φ as the "regular" parameter. The reference prior in the multi-parameter nonregular case discussed in Definition 4, is obtained by maximizing the third term in the asymptotic expansion of Lindley's measure of information with respect to $\pi(\eta, \varphi)$,

$$\log \frac{T}{e} + \frac{k}{2} \log \frac{T}{2\pi e} + \int \pi(\eta, \varphi) \frac{|c(\eta, \varphi)| \sqrt{\det \mathbf{J}_Z^{\varphi\varphi}(\eta, \varphi)}}{\pi(\eta, \varphi)} d\eta d\varphi + o(1), \tag{27}$$

where $c(\eta, \varphi)$ and $\mathbf{J}_Z^{\varphi\varphi}(\eta, \varphi)$ are given in Definition 4. Note that the expression in equation (28) is the counterpart for the asymptotic expansion of Lindley's measure of information for the regular case given in equation (21), and the one-parameter nonregular case given in equation (22).

From the discussion above, several points are worth noting.

First, neither the reference prior $\pi^z_{ref}(\theta)$ nor Jeffreys' prior need be proper; it is just a positive function or a tool for deriving the reference posterior distribution via Bayes' theorem. It is *only the* reference posterior distributions that have a probabilistic interpretation (Bernardo and Smith, 1994, p. 306).

Second, from Definition 2, 3 and 4, and the exposition above, in *general*, Jeffreys' prior and the reference prior are different when the support of the data depends on the parameters. Ofcourse, the two coincide when standard regularity conditions are satisfied so that the support of the data does not depend on the parameters and there is no partitioning of parameters into "nuisance parameters" and "parameters of interest".

Third, the reference prior, like the Jeffreys' prior, is invariant to the following kind of reparametrization. If (η,φ) are reparametrized to (α, β), where $\alpha = \alpha(\eta)$ and $\beta = \beta(\varphi)$ are one-to-one monotonic functions, then the reference prior, $\pi^*(\alpha, \beta)$, for (α, β) is related to the reference prior for (η,φ) by

$$\pi^*(\alpha, \beta) = \left|\frac{\partial \eta}{\partial \alpha}\right| \left|\det\left(\frac{\partial \varphi}{\partial \beta}\right)\right| \pi\left(\eta(\alpha), \varphi(\beta)\right),$$

where $\eta(\alpha)$ and $\varphi(\beta)$ are the inverse transformations, and $\frac{\partial \varphi}{\partial \beta}$ and $\frac{\partial \eta}{\partial \alpha}$ are the Jacobian of the transformation from $\varphi \to \beta$ and $\eta \to \alpha$, respectively.

5 Reference Analysis for Low-Price, Sealed-Bid Auction

For the low-price, sealed-bid auction described in Section 2, *two* features of the sampling density of the data become important in constructing reference priors for $\theta = (\eta, \varphi)$ where η is the "nonregular" parameter and φ the "regular" parameter.

The first pertains to the "regular parameter". I have proved in the last section that the winning bids are not sufficient for the "regular" parameter φ, or the difference in the Fisher information for φ obtained from all the bids and the winning bids, $\mathbf{J}_{\mathbf{B}}^{\varphi\varphi}(\eta, \varphi) - \mathbf{J}_{\mathbf{W}}^{\varphi\varphi}(\eta, \varphi)$, is a *non-null psd* matrix. As a result ***both*** the reference prior for φ and the likelihood as a function of φ will differ depending on whether data on all bids or just the winning bids across auctions is available. I find that ***the reference prior and the exact reference posterior for the "regular" parameters φ is more dispersed when it is obtained from the winning bids across auctions compared to when it is obtained from all the bids within and across auctions.***

The second result pertains to the "nonregular parameter". In the last section I have shown that the minimum winning bid, W_*, is sufficient for η *given* φ. Hence the ***reference prior for η and the reference posterior for the "nonregular" parameter η* given *φ, whether obtained from all the bids in an auction or just the winning bid, is identical to the reference prior for the "non-regular" parameter η given φ obtained from the minimum winning bid.***[15] The intuition for the reference prior for η given φ being identical in the two cases is simple.[16] The reference prior is obtained by maximizing an asymptotic expansion of Lindley's measure of information given in Definition 1. Since the minimum winning

[15] Identical results hold for δ

[16] A formal proof of this is available in Sareen (199b, p. 60). It uses the necessary and sufficinet conditions under which the sample minimum is sufficient for a scalar "nonregular" parameter due to Huzurbazar (1976, pp. 158, 174-186).

bid across auctions is a sufficient statistic for θ, the asymptotic posterior density function in Lindley's measure of information is the same whether it is based on all the bids or just the minimum winning bid. As a result the reference prior from bids within and across auctions and the minimum winning bid is identical. With the reference prior being identical in the two cases, the exact reference posterior is identical as well.

I now obtain the reference priors for $\theta = [\eta, \varphi]'$. I have reparameterized the density functions in equations (1)-(3) such that the nonregular parameter in Definition 4 is $\eta = \frac{\delta\varphi(n-1)}{\varphi(n-1)-1}$, and the regular parameter is φ. The lower bound of the density function for a bid, winning bid, and the minimum winning bid in (1), (2) and (3), respectively, is $l(\eta) = \eta$.

Example 1a: **Bids proportional to cost, reference prior from bids**
From (1), $c(\eta, \varphi) = l'(\eta) f_b(\eta|\eta, \varphi) = \frac{\varphi}{\eta}$. I obtain $\mathbf{J}_{\mathbf{B}}^{\varphi\varphi}(\eta, \varphi)$ as the element in the second row and column of the Fisher information matrix from bids, $\mathbf{J_B}(\eta, \varphi)$; this is given in equation (C.1), Appendix C. $\sqrt{\det \mathbf{J}_{\mathbf{B}}^{\varphi\varphi}(\eta, \varphi)} = \frac{1}{\varphi}\sqrt{nT + \frac{(nT)^2\varphi^2}{(\varphi m-1)^2}}$, where $m = n - 1$. Then from Definition 4 the reference *prior* for $[\eta, \varphi]$ is

$$\begin{aligned}
\pi^b_{ref}(\eta, \varphi) &\propto \frac{\varphi}{\eta}\left[\frac{1}{\varphi}\sqrt{1 + \frac{Tn\varphi^2}{(\varphi m - 1)^2}}\right], \eta, \varphi > 0, \\
&\propto \left\{\frac{1}{\eta}\right\}\left\{\sqrt{1 + \frac{Tn\varphi^2}{(\varphi m - 1)^2}}\right\}. \qquad (28)
\end{aligned}$$

The exact reference *posterior* for $[\eta, \varphi]$ is

$$\begin{aligned}
\pi^b_{ref}(\eta, \varphi|\mathbf{b}) &= \pi^b_{ref}(\varphi|\mathbf{b})\, \pi^b_{ref}(\eta|\varphi, \mathbf{b}), \quad 0 < \eta < w_*; \\
&\propto \left[\sqrt{1 + \frac{nT\varphi^2}{(\varphi m - 1)^2}}\varphi^{nT} e^{-\varphi\left(\sum\sum \log b^i_j\right)}\right]\left[\eta^{nT\varphi - 1}\right], \\
&\propto \left[\sqrt{1 + \frac{nT\varphi^2}{(\varphi m_j - 1)^2}} GA\left(nT + 1, \left(\sum_{j=1}^{T}\sum_{i=1}^{n} \log b^i_j\right)^{-1}\right)\right] \\
&\quad \left[PF\left(nT\varphi, w_*\right)\right]. \qquad (29)
\end{aligned}$$

The first term in the square brackets of (30) is the kernel of the reference posterior for $\varphi, \pi^b_{ref}(\varphi|\mathbf{b})$. The term in the second square brackets is the posterior kernel of $\eta|\varphi$; it belongs to the power-function distribution.

Example 1b: **Reference prior from winning bids**
From equation (17), $c(\eta, \varphi) = l'(\eta) f_w(\eta|\eta, \varphi) = \frac{n\varphi}{\eta}$. I obtain $\mathbf{J}_{\mathbf{W}}^{\varphi\varphi}(\eta, \varphi)$ as the element in the second row and column of the Fisher information matrix from the winning bids, $\mathbf{J_W}(\eta, \varphi)$; this is given in equation (C.2), Appendix C. $\sqrt{\det \mathbf{J}_{\mathbf{W}}^{\varphi\varphi}(\eta, \varphi)} =$

$\frac{1}{\varphi}\sqrt{1+\frac{Tn^2\varphi^2}{(\varphi m-1)^2}}$, where $m = n-1$. Then from Definition 4 the reference *prior* for $[\eta,\varphi]$ is

$$
\begin{aligned}
\pi^w_{ref}(\eta,\varphi) &\propto \frac{\varphi}{\eta}\left[\frac{1}{\varphi}\sqrt{1+\frac{Tn^2\varphi^2}{(\varphi m-1)^2}}\right], \eta,\varphi>0; \\
&\propto \left\{\frac{1}{\eta}\right\}\left\{\sqrt{1+\frac{Tn^2\varphi^2}{(\varphi m-1)^2}}\right\}. \qquad (30)
\end{aligned}
$$

The exact reference *posterior* for $[\eta,\varphi]$ is

$$
\begin{aligned}
\pi^w_{ref}(\eta,\varphi|\mathbf{w}) &= \pi^w_{ref}(\varphi|\mathbf{w})\,\pi^w_{ref}(\eta|\varphi,\mathbf{w}), \quad 0<\eta<w_*; \\
&\propto \left[\sqrt{1+\frac{Tn^2\varphi^2}{(\varphi m-1)^2}}\varphi^T e^{-\varphi n\left(\sum\log w_j\right)}\right]\left[\eta^{nT\varphi-1}\right], \\
&\propto \left[\sqrt{1+\frac{Tn^2\varphi^2}{(\varphi m-1)^2}}GA\left(T+1,\left(n\sum_{j=1}^{T}\log w_j\right)^{-1}\right)\right] \\
&\quad \left[PF\left(nT\varphi,w_*\right)\right]. \qquad (31)
\end{aligned}
$$

The first term in the square brackets of (32) is the kernel of the reference posterior for $\varphi, \pi^b_{ref}(\varphi|\mathbf{w})$. The term in the second square brackets is the posterior kernel of $\eta|\varphi$.

There are several points worth noting in Example 1. First, from equations (29) and (31), η and φ are independent according to the reference prior; this follows from the sampling expectation of the score function, $c(\eta,\varphi)$, and $\sqrt{\det \mathbf{J}^{\varphi\varphi}_{\mathbf{Z}}(\eta,\varphi)}$, $\mathbf{Z}=\mathbf{B},\mathbf{W},W_*$, factorizing into functions of η alone and φ alone.

Second, both the reference prior and the reference posterior for $\eta|\varphi$, whether they are obtained from all the bids or just the winning bids across auctions, are ***identical*** to the reference prior and posterior for $\eta|\varphi$ obtained from the minimum winning bid; this follows from the conditional sufficiency of the minimum winning bid for $\eta|\varphi$. From equations (29) and (31), I find that the reference prior for η, obtained from the bids and the winning bids is *identical* and improper. The reference prior is given by the first term in the curly brackets in the two equations; it is proportional to $(\eta)^{-1}$. The reference prior for $\eta|\varphi$ from the minimum winning bid is the same as well.[17] The

[17] From equation (3), $c(\eta,\varphi)=\frac{nT\varphi}{\eta}$ and $\sqrt{\det \mathbf{J}^{\varphi\varphi}_{\mathbf{W}_*}(\eta,\varphi)}=\frac{1}{\varphi}\sqrt{1+\frac{(nT)^2\varphi^2}{(\varphi m-1)^2}}$ from equation (C.3), Appendix C. The reference *prior* for (η,φ) obtained from the minimum winning bid is

$$\pi^{w_*}_{ref}(\eta,\varphi)\propto\left\{\frac{1}{\eta}\right\}\left\{\sqrt{1+\frac{(nT)^2\varphi^2}{(\varphi m-1)^2}}\right\}.$$

The reference prior for η is $(\eta)^{-1}$.

reference posterior for $\eta|\varphi$ obtained from the bids and the winning bids is given by the second term in the square brackets in equations (30) and (32), respectively. It is identical, belonging to the family of power-function distribution; it is the same as the reference posterior for $\eta|\varphi$ obtained from the minimum winning bid, $\pi^{w_*}_{ref}(\eta|\varphi, w_*)$.[18]

Third, the reference prior for φ from all the bids and the winning bids is ***improper***; it is given by the second term in the curly brackets in equations (29) and (31), respectively. The impropriety of the reference prior in the two cases follows from $\lim_{\varphi\to\infty}\sqrt{1+\frac{Tn\varphi^2}{(\varphi m-1)^2}} = \sqrt{1+\frac{Tn}{m^2}}$ and $\lim_{\varphi\to\infty}\sqrt{1+\frac{Tn^2\varphi^2}{(\varphi m-1)^2}} = \sqrt{1+\frac{Tn^2}{m^2}}$, so that the kernel of the reference prior is *not* integrable. I have plotted the kernel of the reference prior for φ obtained from the bids, $\sqrt{1+\frac{Tn\varphi^2}{(\varphi m-1)^2}}$, and the winning bids, $\sqrt{1+\frac{Tn^2\varphi^2}{(\varphi m-1)^2}}$, in Figure 1, to make this point obvious.

Another feature of the reference prior for φ that I observe from Figure 1 is that most of the mass of the kernel of the reference prior for φ, whether obtained from the bids or the winning bids, is in its right tail. Further, the right tail of the reference prior kernel for φ obtained from the bids is proportional to the right tail of the reference prior kernel for φ from the winning bids.[19] Since reference priors are defined up to a constant of proportionality the right tails of the two reference priors will get absorbed in the constant of proportionality. Hence as $\varphi \to \infty$ it is *immaterial* whether the reference prior is obtained from all the bids or just the winning bids. To illustrate this point, I have normalized the improper reference prior in Figures 2 and 3 to make it proper and plotted the density function of normalized proper prior by truncating φ at $\varphi = 20, 100$.[20] As I increase the truncation from 20 to 100, I find that the normalized prior density function shifts mass towards the right tail in an effort to make the improper reference prior proper. Since the right tail of the normalized reference prior whether obtained from all the bids or just the winning bids almost coincide as I increase the truncation point for φ, the normalized reference prior density functions,

---

[18] I have obtained the reference *prior* for (η, φ) from the minimum winning bid in the last footnote. The reference *posterior* is

$$\pi^{w_*}_{ref}(\eta, \varphi|w_*) \quad \propto \quad \underset{[PF(nT\varphi, w_*)]}{\left[\sqrt{1+\frac{(nT)^2\varphi^2}{(\varphi m-1)^2}}GA\left(2, (nT\log w_j)^{-1}\right)\right]},$$

The second term is the kernel of the reference posterior for $\eta|\varphi$.

[19] Note that the latter is just a scaled-up version of the former where the constant by which it is scaled is n. From the expressions for the reference prior for φ given in the second curly brackets in equations (29) and (31) as $\varphi \to \infty$, the ratio $\frac{\varphi^2}{(\varphi m-1)^2} \to 1$. Hence the right tail behaviour of the reference priors is just a function of n.

[20] This is also the reference prior for φ that emerges from maximizing Lindley's measure of information in equation (27) with respect to $\pi(\eta, \varphi)$ when $\varphi \in [0, 20]$ or $[0, 100]$. The fact that φ is bounded ensures that this reference prior is proper.

whether obtained from the bids or the winning bids will be similar. ***Large*** values of φ are however ***uninteresting*** for the auctions that I am studying; for example, the MLE of φ that I obtain from my data set on winning bids is 0.53. Johnson, Kotz and Balakrishnan (1994, Vol. 2, p. 579) point out that φ can be interpreted as the elasticity of the distribution function of the cost, $F_c(\tau|\eta,\varphi)$, with respect to τ, where τ is a *realized* cost.[21] Since bids are proportional to cost, if cost increases by 10%, φ is the answer to the question, if bids increase by 10% what is the probability that the cost had increased by 10% ? The answer is clearly less than one since bidders submit bids that are padded above cost.

Fourth, the reference prior, as well as, the *exact* reference posterior for φ obtained from the winning bids are ***more dispersed*** compared to when they are obtained from all the bids. From the plot of the reference prior kernel for φ in Figure 1, I observe that the reference prior from the winning bids lies above that obtained from the bids and the right tail of the kernel for φ obtained from the winning bids is thicker than the right tail of the kernel from all the bids. The exact reference posterior for φ obtained from all the bids and the winning bids, respectively, is given by the first term in the square brackets in equations (30) and (32). The first term in the posterior for φ is contributed by the reference prior for φ; this has been plotted in Figure 1. The second term is a Gamma density function. This density function has fatter tails for the winning bids than all the bids since the winning bid is the *lowest* bid in an auction j, so that $\sum_{j=1}^{T}\sum_{i=1}^{n}\log b_j^i > n\sum_j = 1^T \log w_j$, and $nT+1 > T+1$. These features of the reference prior for φ follows from the winning bids not being sufficient for the regular parameter φ, or all the bids containing *strictly* more Fisher information than the winning bids for the regular parameter φ.

I illustrate the difference in inference about φ from data on all bids compared with data on winning bids by considering the low-price, sealed-bid auctions described in Section 4 assuming that the number of potential bidders in each auction is six, $n = 6$.[22] I *do not* have data on bids for individual auctions. Hence I use an *approximation* for $\sum_{j=1}^{T}\sum_{i=1}^{n}\log b_j^i$ in equation (30).[23] The marginal posterior for φ is given by the

[21]If a variable $Z_j^i \sim PA(\eta,\varphi)$, then φ can be interpreted as the following elasticity,

$$\varphi = \frac{d\log F_z(z|\eta,\varphi)}{d\log z},$$

where $F_z(z|\eta,\varphi)$ is the distribution function of Z_j^i.

[22]The average number of participants in the auctions conducted by IOC is 6.

[23]Using the MLE of η and φ, I draw five bids above the winning bid in an auction, from the Pareto distribution $PA(\widehat{\eta^{mle}},\widehat{\varphi^{mle}})$. $\widehat{\eta^{mle}} = 95.51$ U.S. dollar per metric tonne, is the minimum winning bid in my data set. $\widehat{\varphi^{mle}} = 0.53$ and has been obtained as follows

$$\widehat{\varphi^{mle}} = T\sum_{j=1}^{n} n_j \log\left(\frac{w_j}{\widehat{\eta^{mle}}}\right),$$

where $T = 37$ is the number of auctions on which data is available. I do this for each of the $T = 37$

term in square brackets in equations (30) and (32) from all the bids in the auctions and the winning bids, respectively. The first term (square-root term) in the square brackets in both equations is the contribution of the reference prior and the second term ($GA(\bullet)$) the contribution of the likelihood function to the marginal posterior for φ. I have plotted the kernel of the marginal reference posterior for φ given by the term in the square brackets in equation (30) and (32) in Figure 4 and 5, respectively. The kernel of the Gamma density functions, the contribution of the likelihood function to the posterior for φ, has also been plotted. While the mode of the two posterior distributions is the same, their tails are different. The posterior for φ obtained from the winning bids in Figure 5 has fatter tails than the posterior for φ based on all the bids in Figure 4. As a result interval estimates for φ would be different; specifically, the highest posterior density interval based on the winning bids will be ***wider*** compared with that based on all the bids. Point estimates for φ would be different if the quadratic loss function is used since the posterior mean is different between Figures 4 and 5, respectively. One scenario were ***same*** point estimates would be obtained was if the ***all-or-nothing*** loss function was used; the mode of the two posterior distributions in Figures 4 and 5 is the same.

From the discussion above the message to an empirical researcher would be to obtain the data on bids and base inference on the posterior obtained from all the bids. It could be the case that she is ***unable*** to obtain data on the losing bids so that she has to work with winning bids only. In this case inference about φ should be based on the posterior that is the product of the likelihood of the winning bids and the reference prior for φ based on all bids; this has been labelled "Bposterior" in Figure 5. I show that this posterior has ***thinner tails*** than the posterior that is the product of the likelihood of the winning bids and the reference prior for φ based on winning bids; the latter is labelled "Wposterior" in Figure 5.

Finally, like Arnold and Press (1983, pp. 192-193), the exact reference posterior for $\eta|\varphi$ belongs to the power-function distribution. They arrive at a noninformative prior as a special case of the conjugate prior for the scale parameter of a Pareto likelihood when the shape parameter is known; since this noninformative prior is the same as the reference prior for η given φ, that I have derived above, I obtain a power-function distribution for η given φ as well.

However Arnold and Press (1982, pp. 293-297) obtain an improper bivariate posterior for (η, φ) with a Pareto likelihood and independent, noninformative priors for the two parameters. The reason for this was that the likelihood function becomes unbounded at $\eta = 0$; with both the noninformative prior for η and the likelihood putting an infinite mass at $\eta = 0$, the marginal posterior density function for η becomes unbounded for η near 0.[24] As a result they recommend that this noninformative prior

auctions and obtain $\sum_{j=1}^{T}\sum_{i=1}^{n}\log b_j^i$.

[24] Arnold and Press assume that the two parameters are *apriori* independent with η having a power function distribution, $PF(a, b)$ and φ a Gamma distribution, $GA(c, d)$. Their noninformative prior emerges by setting $c = d = 0$ and $a = 0$, $b = \infty$; $a = 0$, $b = \infty$ leads to the asymptote in the

not be used to represent diffuseness. I do not observe the kind of anomalous behavior noted by Arnold and Press when I use the reference prior to represent diffuseness. From equations (30) or (32), the posterior kernel of (η, φ) would become unbounded near $\eta = 0$ if $0 \leq nT\varphi < 1$; however, $0 \leq nT\varphi < 1$ violates the positive support of the density of $\eta|\varphi$ since $\eta = \frac{\delta\varphi(n-1)}{\varphi(n-1)-1}$. Thus even though the likelihood function of the bids or the winning bids given in equations (15) and (16) respectively, puts an infinite mass near $\eta = 0$, the reference prior puts no mass near $\eta = 0$. Hence, I do not observe the impropriety in the posterior for (η, φ) near $\eta = 0$ like Arnold and Press. The message is that it is attractive to use the *reference prior for* (η, φ) *to represent diffuseness* in the Pareto case since the exact reference posterior for (η, φ) is proper.

6 Reference Analysis for Job-Search Model

In this Section I discuss reference analysis for the reserve wage ξ, the "nonregular" parameter in job-search models. *Two* features of the sampling density of the data are relevant in constructing the reference prior for ξ.

First, *unlike* the structural auction model described in this paper the *observed* data on the accepted wage offers and duration of unemployment is *not sufficient* for ξ.

Second the support of the data under *partial observability* differs from that under *full observability*. Under partial observability, the accepted wage offer of an individual is greater than the reserve wage, ξ; under full observability all the rejected wage offers of an individual are less than ξ, and the accepted wage offer greater than ξ.

The reference prior for the "nonregular" parameter is given in Definition 3. I *cannot* apply Definition 3 to obtain the reference prior for ξ in the job-search model. The reason is that I am no longer in the univariate *iid* setup under which the reference prior is obtained in Definition 3;[25] Hence I propose a *noninformative* prior for ξ for the job-search model and provide the *link* between this proposed prior and the reference prior of Ghosal and Samanta (1997) given in Definition 3. **Like** Definition 3, the reference prior for ξ under *full observability* is proportional to $|c(\xi)|$. $c(\xi)$ is the sampling expectation of the score function of the density function given by equation (4) under *full observability* and by (5) under *partial observability*. **Unlike** Definition 3 it does not simplify to the expression given by equation (23) since the accepted wage offer is greater than ξ and the rejected wage offers less than ξ.

posterior for (η, φ) in the direction of η near $\eta = 0$.

[25] In personal communication S. Ghosal has pointed out to me the problem with the job search setup. The reference prior is obtained by maximizing the asymptotic expansion of the expected Kullback-Leibler divergence. The properties of this expansion are unproven in the literature for the multivariate, non-*iid* case like the job search model. I could condition on $(\mathbf{S}, \mathbf{T})$ since their distributions are "regular"; I would then be in a univariate setup since the wage offers would be the relevant variable. However the accpeted wage offer and the rejecetd wage offers are neither identically, nor independently, distibuted. I have noted earlier that the accepted wage offers are greater than ξ and the rejected wage offers less than ξ.

I find that the *proposed* noninformative prior for ξ ***under partial observability has a relatively flat left tail unlike the proposed noninformative prior for*** ξ ***under full observability; the latter is bimodal, with one of the modes being in the left tail***. In job-search models the missing data on rejected wage offers and the number of such offers is *unavailable.* As a result, comparing the posterior for ξ under *partial observability* and *full observability* is uninteresting. I compare instead the posterior for ξ under *partial observability* with the posterior for ξ that is the product of the likelihood under *partial observability* and the *proposed* prior under *full observability.* Both posteriors are the relevant priors restricted to the interval $\xi \geq w_*$, the minimum accepted wage offer across individuals. Like the reference priors for ξ under *full* and *partial observability*, the ***latter posterior for*** ξ ***gives more information in the left tail compared to the former posterior for*** ξ**.**

I now obtain the *proposed* noninformative prior for ξ under *full observability* and *partial observability* assuming that φ, the "regular" parameters, are given.

Example 2a: **Reference analysis under "full observability"**

Under *full observability* the *proposed* noninformative *prior* for ξ given φ is

$$\begin{aligned} \pi^{suwt}(\xi|\varphi) &\propto |c_{suwt}(\xi)|, \\ &\propto \left| \frac{fw(\xi|\varphi)}{\bar{F_w}\ (\xi|\varphi)} - \frac{fw(\xi|\varphi)}{F_w(\xi|\varphi)} \right|; \end{aligned} \tag{32}$$

this has been derived in Appendix E. $c_{suwt}(\xi)$ is the sampling expectation of the score of the density function given by (4),

$$c_{suwt}(\xi) = E_{S\mathbf{U}WT|\theta} \left(\frac{\partial \log f_{\mathbf{suwt}}(s_j, \mathbf{u}_j, w_j, t_j|\eta, \varphi)}{\partial \eta} \right).$$

The likelihood function of the N agents is the product, over these agents, of the density function given by (4). The exact *posterior* for ξ is proportional to

$$\pi^{suwt}(\xi|\varphi, \mathbf{u}, \mathbf{w}, \mathbf{s}, \mathbf{t}) \propto \pi^{uwst}(\xi|\varphi) \prod_{j=1}^{N} \left[\lambda^{s_j+1} e^{-\lambda t_j} fw(w_j|\mu, \sigma) \prod_{i=1}^{s_j} fw(u_j^i|\mu, \sigma) \right],$$

$$u_* \leq \xi < w_*;$$

$$\pi^{suwt}(\xi|\varphi, \mathbf{u}, \mathbf{w}, \mathbf{s}, \mathbf{t}) \propto \pi^{uwst}(\xi|\varphi), \qquad u_* \leq \xi < w_*. \tag{33}$$

The *posterior* for ξ is the noninformative prior given by (33) restricted to the interval $[u_*, w_*]$. In the simulated data set used by Lancaster (1997, p. 167), this interval is $[12.12, 12.25]$. Since this is a short interval, inference about ξ based on this posterior is very accurate. Lancaster points out that this short interval is the result of the superconsistency of U_* and W_* for ξ.

Example 2b: Reference analysis under "partial observability"

The density function of the accepted wage offers and the duration of unemployment under *partial observability* is given by equation (5). I have obtained $c_{wt}(\xi)$ in Appendix E. It is the sampling expectation of the score of the density function given by (5),

$$c_{wt}(\xi) = E_{WT|\theta}\left(\frac{\partial \log f_{wt}(w_j, t_j|\eta, \varphi)}{\partial \eta}\right).$$

The *proposed* noninformative *prior* for ξ is

$$\begin{aligned} \pi^{wt}(\xi|\varphi) &\propto |c_{wt}(\xi)|, \\ &\propto \frac{fw(\xi|\varphi)}{\bar{F_w}\ (\xi|\varphi)}. \end{aligned} \tag{34}$$

The *posterior* for ξ is

$$\pi^{wt}(\xi|\varphi, \mathbf{w}, \mathbf{t}) \quad \propto \quad \pi^{wt}(\xi|$$

I now discuss the link between the *proposed* noninformative priors and the reference prior of Ghosal and Samanta (1997) given in Definition 3.

First, the *proposed* noninformative prior in (35) under *partial observability* is **identical** to the *reference* prior given in Definition 3 when it is obtained from the sampling density of the accepted wage offers conditional on the duration of unemployment, $W_j|T_j$ instead of the *joint* sampling density of (W_j, T_j) as in Example 2b. It seems reasonable to condition on T_j, the duration of unemployment of agent j, since the support of the density of T_j, *does not* depend on any parameter. From Lancaster (1997, p. 167) the density function of $W_j|T_j$ is

$$f_{w|t}(w_j|T_j, \xi, \varphi) = \frac{fw(w_j|\varphi)}{\bar{F_w}\ (\xi|\varphi)}, \qquad w_j \geq \xi. \tag{36}$$

This density function satisfies the assumptions of Definition 3. From Definition 3, the *reference prior* for ξ under *partial observability* assuming that the duration of unemployment of an agent, t_j, is given,[26] is

$$\pi^{w|t}_{ref}(\xi|\varphi) \propto \frac{fw(\xi|\varphi)}{\bar{F_w}\ (\xi|\varphi)}. \tag{37}$$

The reference prior in (38) is **identical** to the proposed *noninformative* prior in (35) under *partial observability*.[27]

[26] In accordance with Definition 3, this reference prior for ξ given φ is proportional to $c_{w|t}(\xi)$, the sampling expectation of the score of the density function of $W_j|T_j$ given by (37). That is,

$$c_{w|t}(\xi) = E_{W|T,\theta}\left(\frac{\partial \log f_{w|t}(w_j|t_j, \eta, \varphi)}{\partial \eta}\right) = \frac{fw(\xi|\varphi)}{\bar{F_w}\ (\xi|\varphi)}.$$

[27] Using this prior, the exact *reference posterior* for ξ is

$$\pi^{w|t}_{ref}(\xi|\varphi, \mathbf{w}, \mathbf{t}) \propto \pi^{w|t}_{ref}(\xi|\varphi)\left[\frac{1}{\bar{F_w}\ (\xi|\varphi)}\right]^N, \xi \leq w_*.$$

Second, the *proposed* noninformative prior for ξ under *full observability* given by (33) is the sum of two terms; one of the terms is the contribution of the observed data and the other of the missing data.

The first term is the *proposed* noninformative prior for ξ given φ under *partial observability.* I have just indicated that this is **identical** to the reference prior for ξ given φ obtained from the conditional sampling density $W_j|T_j$ **instead** of the joint density of (W_j, T_j) as in Example 2b under partial observability.

The second term in the proposed noninformative prior for ξ under *full observability* is the contribution of the **missing** data to the reference prior for ξ given φ. The rejected wage offers of an agent and the number of these rejected offers, $(\mathbf{U}_j, S_j)$, comprises the missing data of an agent j. The second term is the reference prior for ξ given φ obtained from the sampling density of the rejected wage offers conditional on the number of such offers, $\mathbf{U}_j|s_j$. From Lancaster (1997, p. 166) this conditional density is

$$f_{u|s}(\mathbf{u}_j|s_j,\xi,\varphi) = \prod_{i=1}^{s_j} \frac{fw(u_j^i|\mu,\sigma)}{F_w(\xi|\varphi)}, \qquad u_j^1, ..., u_j^{s_j} < \xi. \tag{38}$$

Like the *partial observability* scenario conditioning on S_j is reasonable since the density function of S_j is "regular"; that is, the support of this density function does not depend on any parameter. Since this density function satisfies the assumptions under which the reference prior is defined in Definition 3, the *reference prior* for ξ given φ,[28] from the **missing** data is

$$\pi_{ref}^{u|s}(\xi|\varphi) \propto \frac{fw(\xi|\varphi)}{F_w(\xi|\varphi)}. \tag{39}$$

I have plotted this reference prior in Figure 6. Most of the mass of this prior is in the interval $[0, 4]$; outside this interval the prior is relatively flat. In general, this prior is a decreasing function of ξ.

The point to note is that this reference prior for ξ given φ from missing data and that obtained under *partial observability* in Example 2b are putting mass on **different** parts of the parameter space for ξ. This is to be expected since reference priors reflect features of the sampling density of the data. As mentioned before the support of the density functions of the accepted wage offers and the rejected offers is different. While the accepted wage offer is greater than ξ, the rejected wage offers are less than ξ. Under *full observability*, information about the **left** tail of the

With the term in the square brackets being a strictly increasing function of ξ, this posterior is increasing in ξ like Example 2b; again, it puts the maximum mass at $\xi = w_*$.

[28] Again from Definition 3, this reference prior for ξ given φ is proportional to $c_{u|s}(\xi)$, the sampling expectation of the score of the density function of $U_j|S_j$ given by (40). That is,

$$c_{u|s}(\xi) = E_{U|S,\theta}\left(\frac{\partial \log f_{u|s}(u_j^i|s_j,\eta,\varphi)}{\partial \eta}\right) = \frac{fw(\xi|\varphi)}{F_w(\xi|\varphi)}.$$

proposed prior for ξ is provided by the missing data on rejected wage offers of an agent; information about the **right** tail is provided by the accepted wage offer of the agent. Hence the *proposed* noninformative prior for ξ given φ under *full observability* resembles the reference prior for ξ given φ under *partial observability* in that part of the parameter space for ξ where this prior puts most of its mass. In that part of the parameter space for ξ where the reference prior for ξ given φ obtained from missing data puts most of it's mass, it resembles this prior.

Third, comparing inference for ξ given φ under *full observability* with that under *partial observability* is an uninteresting exercise in job-search models. The missing data on the rejected wage offers and the number of such offers is **never** observed by the econometrician. The **interesting** comparison here is between the posterior for ξ given φ under *partial observability* and the posterior for ξ given φ that is the product of the likelihood under *partial observability* and the reference prior under *full observability*. The latter posterior is the proposed noninformative prior given by equation (33) in Example 2a restricted to the interval $0 < \xi \leq w_*$; $w_* = 12.24$ in the simulated data set used by Lancaster. The posterior for ξ given φ under *partial observability* is given by equation (36) in Example 2b. While the two posteriors are identical in the interval $[2, w_*]$, the latter is flat in the left tail. Hence the posterior for ξ given φ that is based on the *proposed* prior under *full observability* gives information about ξ in the **left** tail **unlike** the posterior for ξ given φ under *partial observability*. One scenario where inference for ξ given φ would be similar is if it is based on the all-or-nothing loss function since both posteriors put maximum mass at $\xi = w_* = 12.24$.[29]

7 Reference Analysis for Roy Selection Model

In this Section reference analysis for the coefficient of correlation between the potential earnings of an individual in the two sectors of choice, ρ_{12}, is discussed. Since the support of the data does not depend on parameters in the Roy model and I am not partitioning parameters into "nuisance parameters" and "parameters of interest", the reference prior and Jeffreys' prior are identical.

Two features of the sampling density of the data are relevant for reference analysis of ρ_{12}.

First, ρ_{12} does not enter the likelihood function *explicitly* under *partial observability*. However data under partial observability contains expected information about ρ_{12}. The positive definiteness of the covariance matrix $\mathbf{\Omega}$ given in equation (8) *restricts* the support of ρ_{12}. This restricted support depends on the parameters ρ_{1u} and ρ_{2u} that are *identified* under *partial observability*. This is the **channel** through which expected information is obtained about ρ_{12} from the observed data on realized earnings in the two sectors and the sectoral choice of agents.

[29]Lancaster (1997, pp. 168, 171-172) notes a similar point without making any parametric assumption about the prior for ξ.

Second, if the missing data on potential earnings of an individual in the sector that she did not chose was available as well, then the expected information about ρ_{12} under *full observability* is strictly greater than that under *partial observability.*

Like the job-search models, missing data is *unavailable* in the Roy model. As a result it is *not* interesting to compare reference analysis for ρ_{12} under *partial observability* with that under *full observability.* Hence I confine myself to the likelihood of the data under *partial observability.* I compare the posterior for ρ_{12} that is a product of this likelihood and the reference prior under *full observability* with the reference posterior for ρ_{12} under *partial observability.*

The Jeffreys' or reference prior for ρ_{12} under *partial observability* is not defined since ρ_{12} does not explicitly enter the sampling density under *partial observability.* Hence following *convention,* under *partial observability,* I use a proper Uniform prior for ρ_{12} conditional on the identified parameters, ρ_{1u} and ρ_{2u}.[30] The support of this prior is the interval to which ρ_{12} is restricted by the positive definiteness condition of the covariance matrix; this is given in (19).

As long as the prior for the identified parameters is proper, marginal inference about ρ_{12} ***is possible irrespective of the prior used.***[31] That is, the marginal prior and posterior for ρ_{12} are different indicating learning from data about ρ_{12}. ***Inference about*** ρ_{12} ***conditional on the identified parameters,*** ρ_{1u} ***and*** ρ_{2u}***, is possible only if the prior used is the Jeffreys' prior under full observability.***[32] Conditional inference under *partial observability* is not feasible since the prior and posterior for $\rho_{12}|\rho_{1u}, \rho_{2u}$ are identical. This follows from the prior and the likelihood incorporating **identical** information about ρ_{12}; both restrict the support of ρ_{12} through the positive definiteness of $\mathbf{\Omega}$. In effect, an ***additional*** channel of learning is introduced *via* the prior when inference is based on the conditional posterior of ρ_{12} that is a product of the likelihood under *partial observability* and the reference prior under *full observability.*

In the discussion below it is assumed that $\varphi_* = [\gamma, \beta_1, \beta_2, \sigma_1, \sigma_2]'$ is given. Marginal inference about ρ_{12} indicates inference about ρ_{12} conditional on φ_*. Conditional

[30]Vijverberg (1993, p. 79) specifies this prior for ρ_{12} in the Roy model. Poirier (1998, p. 493) follows this convention when discussing a special case of the censored sampling model of Manski (1994). In this class of model, like the Roy model, the nonidentified parameter does not enter the likelihood function of the observed data explicitly. The nonidentified parameter has a restricted support which is a function of the identified parameters.

[31]Marginal inference about ρ_{12} is an interesting exercise 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 charcteristics, such as punctuality and dilligence, enhance an agent's productivity, they will do so irrespective of the sector chosen by this agent.

[32]When interest is in correlation between the unobserved factors affecting productivity in the two sectors conditional on the self-selection mechanism, then conditional inference about ρ_{12} is the way to go. For example, the posterior for ρ_{12} conditional on $\rho_{1u} > 0, \rho_{2u} < 0$ would answer the question "Given that an agent has chosen sector 1 or 2, is it possible that this agent can be more productive on an average in both sectors ?". See Vijverberg (1993, pp. 75-76) for further details.

inference will refer to inference about ρ_{12} conditional on ρ_{1u}, ρ_{2u} and φ_*. Conditioning on φ_* will be suppressed for notational convenience.

Example 3a: Reference analysis under "full observability"
Given ρ_{1u}, ρ_{2u} and φ_*, the Fisher information about ρ_{12} under full observability has been obtained in Appendix F. The square-root of this Fisher information is the Jeffreys' prior or the reference prior for ρ_{12} conditional on ρ_{1u}, ρ_{2u} and φ_*. It is given by

$$\pi^{I^* Y_1^* Y_2^*}(\rho_{12}|\rho_{1u},\rho_{2u}) \propto \sqrt{\frac{1}{D}\left[1+\frac{2}{D}(\rho_{1u}\rho_{2u}-\rho_{12})^2\right]}, \quad -1 \le \rho_{12} \le 1; \tag{40}$$

D is the determinant of the covariance matrix $\mathbf{\Omega}$ given in equation (8),

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

Since the potential earnings of an agent in the sector she did not choose are **never** observed by the econometrician, obtaining the posterior for ρ_{12} under *full observability* is an uninteresting exercise. I instead obtain the posterior for ρ_{12} conditional on ρ_{1u}, ρ_{2u} and φ_* that is a product of the Jeffreys' prior given by (42) and the likelihood under *partial observability.* The likelihood under partial observability is given by equation (12). ρ_{12} does not enter this likelihood function. I have mentioned before that the positive definiteness of $\mathbf{\Omega}$ constraints the range of ρ_{12} to $\Psi(\rho_{1u},\rho_{2u})$; this is given in equation (19). Taking this constraint into account I rewrite the likelihood function in equation (12) as

$$L(\rho_{12},\rho_{1u},\rho_{2u};\mathbf{Y}_1^O,\mathbf{Y}_2^O,\mathbf{I}) = \left\{\prod_{j=1}^{n}\left[\int_0^\infty f_{1u}(\bullet)\right]^{I_j}\left[\int_{-\infty}^0 f_{2u}(\bullet)\right]^{1-I_j}\right\}\mathbf{1}_{\Psi(\rho_{1u},\rho_{2u})}(\rho_{12}); \tag{41}$$

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

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

The posterior for ρ_{12} conditional on ρ_{1u}, ρ_{2u} and φ_* is a product of this likelihood and the Jeffreys' prior given by (42),

$$\pi^{I^* Y_1^* Y_2^*}(\rho_{12}|\rho_{1u},\rho_{2u},\mathbf{Y}_1^O,\mathbf{Y}_2^O,\mathbf{I}) \propto \pi^{I^* Y_1^* Y_2^*}(\rho_{12}|\rho_{1u},\rho_{2u})\mathbf{1}_{\Psi(\rho_{1u},\rho_{2u})}(\rho_{12}). \tag{42}$$

Thus, the posterior is just the Jeffreys' prior for ρ_{12} **restricted** to the interval $\Psi(\rho_{1u},\rho_{2u}) = [\rho_{12}^l,\rho_{12}^u]$.

Example 3b: Reference analysis under "partial observability"
Since ρ_{12} does not explicitly enter the likelihood under *partial observability* given by equation (12) or (43), the Fisher information for ρ_{12}, and as a result, Jeffreys' prior for ρ_{12} given ρ_{1u}, ρ_{2u} and φ_* is not defined. Following convention (Vijverberg 1993,

Poirier 1998) a proper, uniform prior defined on the support $\Psi(\rho_{1u}, \rho_{2u})$ is considered; thus

$$\pi^{IY_1^oY_2^o}(\rho_{12}|\rho_{1u}, \rho_{2u}) = \frac{1}{\rho_{12}^u - \rho_{12}^l}, \rho_{12}^l \leq \rho_{12} \leq \rho_{12}^u. \tag{43}$$

The likelihood under *partial observability* is given by (43); ρ_{12} enters the likelihood only through the support $\Psi(\rho_{1u}, \rho_{2u})$. This restricted support of ρ_{12} has already been taken into account in constructing the prior for $\rho_{12}|\rho_{1u}, \rho_{2u}$ in equation (45). Hence the **posterior** and **prior** for ρ_{12} given ρ_{1u}, ρ_{2u} and φ_* are **identical** under *partial observability.*

The following point are worth noting when conducting inference for ρ_{12} conditional on ρ_{1u}, ρ_{2u} and φ_*.

First, under *partial observability* the prior for $\rho_{12}|\rho_{1u}, \rho_{2u}, \varphi_*$ given by (45) and the likelihood function given by (44) are incorporating **identical** information about ρ_{12}. This information is obtained from the identified parameters, ρ_{1u}, ρ_{2u}, *via* the positive definiteness of the covariance matrix. Hence the prior and the posterior for $\rho_{12}|\rho_{1u}, \rho_{2u}, \varphi_*$ are identical. This phenomenon of the prior and the posterior for for the unidentified parameter conditional on the identified parameters being identical has been described by different phrases. Dawid (1979) calls it the "nonidentifiablity" of ρ_{12}; Poirier (1998) uses the phrase "the data are conditionally uninformative for ρ_{12}". Their argument runs as follows. Since

$$\begin{aligned}\pi^{I^*Y_1^*Y_2^*}(\rho_{12}|\rho_{1u}, \rho_{2u}, \mathbf{Y}_1^O, \mathbf{Y}_2^O, \mathbf{I}) \quad \propto \quad & L(\rho_{12}, \rho_{1u}, \rho_{2u}; \mathbf{Y}_1^O, \mathbf{Y}_2^O, \mathbf{I})\pi(\rho_{1u}, \rho_{2u}) \\ & \pi^{I^*Y_1^*Y_2^*}(\rho_{12}|\rho_{1u}, \rho_{2u}),\end{aligned} \tag{44}$$

the prior and the posterior for $\rho_{12}|\rho_{1u}, \rho_{2u}, \varphi_*$ will be identical if the likelihood is free of ρ_{12}. This is transparent for the likelihood under *partial observability* given by equation (43). With the support, $\Psi(\rho_{1u}, \rho_{2u})$, of ρ_{12} depending only on ρ_{1u}, ρ_{2u}, the likelihood function can be parametrized in terms of the identified parameters *exclusively.*

An **additional** situation for the "nonidentifiability of ρ_{12}" or the "the data being conditionally uninformative for ρ_{12}" follows from (47) in case **noninformative priors** based on the sampling density of the data are being used. If both the prior and the likelihood use the **same channel** to extract information about the nonidentified parameter from the identified parameters, then the prior and the posterior for the nonidentified parameter conditional on the identified parameters will be identical. This is what is happening under *partial observability* when the conventional noninformative prior used by Vijverberg (1993) and Poirier (1998) is used. Both the likelihood and the prior are indicator functions restricting the support of $\rho_{12}|\rho_{1u}, \rho_{2u}, \varphi_*$ to $\Psi(\rho_{1u}, \rho_{2u})$. Thus it is not just the conditioning on the identified parameters but the **way** the **conditioning** is being done that is driving the prior and the posterior for $\rho_{12}|\rho_{1u}, \rho_{2u}, \varphi_*$ to be identical. When inference is based on the posterior that is a product of the likelihood under *partial observability* and the reference prior under *full observability*, the likelihood and prior provide different information about ρ_{12}. The likelihood function provides information about ρ_{12} by restricting it's support; this support is a function of the identified parameters through the positive definiteness of $\mathbf{\Omega}$. In the next point the information provided by the prior for ρ_{12} conditional on $\rho_{1u}, \rho_{2u}, \varphi_*$ is discussed.

Second, Jeffreys' prior for ρ_{12} conditional on $\rho_{1u}, \rho_{2u}, \varphi_*$ under *full observability* provides average information about ρ_{12} from **both** the observed and missing data. Note that ρ_{12} enters the likelihood function under *full observability* **explicitly**; that is, ρ_{12} is identified under full observability. I have plotted this prior given by equation (42) in Figure 7. I find that it is approximately symmetric around the mid-point of the interval $\Psi(\rho_{1u}, \rho_{2u})$.[33] It also has two points of inflection at

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

Further as $\rho_{12} \rightarrow \rho_{12}^l, \rho_{12}^u$, $\pi^{I^*Y_1^*Y_2^*}(\rho_{12}|\rho_{1u}, \rho_{2u}) \rightarrow \infty$ indicating that the prior has two modes at the end-points of the interval $\Psi(\rho_{1u}, \rho_{2u})$. In general the prior puts substantial mass in the tails. What information are these features of the prior conveying? Roy model has empirical content if $\rho_{12} \neq 0$. This is what Jeffreys' prior under *full observability* is ensuring. 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 it's 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 noninformative prior under *partial observability* given by (45) which makes each point on the interval $\Psi(\rho_{1u}, \rho_{2u})$ **equally likely**.

Third, if inference about ρ_{12} conditional on $\rho_{1u}, \rho_{2u}, \varphi_*$ is based on the prior under *full observability* and the likelihood under *partial observability* as in Example 3a, then the issue is whether inference is at all possible? That is, is the posterior $\rho_{12}|\rho_{1u}, \rho_{2u}, \mathbf{Y}_1^O, \mathbf{Y}_2^O, \mathbf{I}$ in (44) proper? The prior for $\rho_{12}|\rho_{1u}, \rho_{2u}, \varphi_*$ under *full observability* is improper; that is, it's integral diverges to infinity. In addition, ρ_{12} does not enter the likelihood function under *partial observability* explicitly. Poirier (1998, pp 504-505) and Ghosh, et al (1999) discuss the problem of a nonidentified parameter combined with an improper prior for this nonidentified parameter.[34] Whether the posterior is proper or not depends on the nature of identifiability and the kind of information provided by the prior even though it is improper. I have shown that

[33] If I normalize the function in (42), I find that $P(-1 \leq \rho_{12} \leq \rho_{12}^m) = P(\psi_m \leq \rho_{12}^m \leq 1) = 0.5$, where ρ_{12}^m is the mid-point of $\Psi(\rho_{1u}, \rho_{2u})$.

[34] Poirier (1998, pp. 504-505) discusses a location-scale model where the location parameter is the sum of two parameters, both of which are not identified individually. Ghosh, et al (1999) discuss item response models with improper prior for the nonidentified parameter. They provide sufficient conditions under which the posterior is proper for certain one-to-one transformation of the nonindentified parameter.

Jeffreys' prior for $\rho_{12}|\rho_{1u}, \rho_{2u}, \varphi_*$ puts most of its mass away from the centre of the interval $\Psi(\rho_{1u}, \rho_{2u})$. If it was possible to bound the mass it puts in the tails of the prior, then the posterior would be proper and inference about ρ_{12} conditional on ρ_{1u}, ρ_{2u} would be possible. This is what the likelihood function accomplishes. The likelihood function restricts $\mathbf{\Omega}$ to be positive definite; a necessary condition for this is $D > 0$; or equivalently in terms of ρ_{12}, $\rho_{12}^l < \rho_{12} < \rho_{12}^u$. This condition bounds the prior kernel of $\rho_{12}|\rho_{1u}, \rho_{2u}, \varphi_*$ in (43) leading to a proper posterior $\rho_{12}|\rho_{1u}, \rho_{2u}, \mathbf{Y}_1^O, \mathbf{Y}_2^O, \mathbf{I}$ in (44); I prove this in Appendix F. Inference about ρ_{12} conditional on the identified parameters is thus feasible if Jeffreys' prior under *full observability* is used.

Finally, marginal inference about ρ_{12} is possible irrespective of the prior used.[35] In view of Poirier (1998, p. 485) this is not surprising under *partial observability* in Example 3b since ρ_{12} and ρ_{1u}, ρ_{2u} are dependent a priori and the noninformative prior $\rho_{12}|\rho_{1u}, \rho_{2u}, \varphi_*$ in (45) is proper.. When Jeffreys' prior for $\rho_{12}|\rho_{1u}, \rho_{2u}, \varphi_*$ is based on the sampling density under *full observability* as in Example 3a, then the feasibility of marginal inference about ρ_{12} follows from the conditional posterior $\rho_{12}|\rho_{1u}, \rho_{2u}, \varphi_*, \mathbf{Y}_1^O, \mathbf{Y}_2^O, \mathbf{I}$ being proper. The marginal posterior for ρ_{12}, $\pi^{I^*Y_1^*Y_2^*}(\rho_{12}|\mathbf{I}, \mathbf{Y}_1^O, \mathbf{Y}_2^O)$, is

$$\int\int \pi^{I^*Y_1^*Y_2^*}(\rho_{12}|\rho_{1u}, \rho_{2u}, \mathbf{I}, \mathbf{Y}_1^O, \mathbf{Y}_2^O) f_{IY_1^OY_2^O}(\mathbf{I}, \mathbf{Y}_1^O, \mathbf{Y}_2^O|\rho_{1u}, \rho_{2u}) d\rho_{1u} d\rho_{2u}.$$

I have mentioned above that when the prior given in (45) is used in conjunction with the likelihood under *partial observability*, the posterior in Example 3a is proper. That is, the first term of the integrand is bounded. The second term in the integrand is the density function of the data given by (43); this is bounded as well. Since the integrand is bounded and the support of ρ_{1u}, ρ_{2u} is $[-1, 1]$, the marginal posterior for ρ_{12} is proper; inference about ρ_{12} is feasible as a result.[36]

8 Conclusion

I quantify the loss in expected information about parameters in terms of the Fisher information for three missing data problems: a structural, low-price, sealed-bid auction, job-search and Roy selection model. The difference in the three models is that while an empirical researcher may be able to obtain the missing data in the low-price, sealed-bid auction discussed, missing data is *unavailable* in the job-search and the Roy selection model.

For these three models, I next link this loss in information to the specification of reference priors based on just the observed data *and* both the observed and missing data. I find that the reference prior based on the sampling density of both the missing and the observed data conveys more information in the tails. As a result I suggest

[35] Marginal inference refers to inference about ρ_{12} "marginal" of the identified parameters ρ_{1u} and ρ_{2u} but conditional on the identified parameters φ_*. I have indicated that φ_* do not enter the support of ρ_{12} *unlike* ρ_{1u} and ρ_{2u}.

[36] Once again the difference in Example 3a and 3b will be in the tails of the marginal posterior for ρ_{12}

that when missing data is *unavailable*, inference about the parameter be based on the posterior that is a product of this prior and the likelihood of the observed data.

In addition, for the Roy model I show that inference about the unidentified coefficient of correlation between the earnings in the sectors of choice conditional on the identified parameters is possible only if the posterior is a product of the likelihood of the observed data and the reference prior under *full observability.*

References

[41] Angrist, J.D., Imbens, G.W. and Rubin, D.B., 1996, Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association* 91, 444-455.

[41] Arnold, B.C. and J. Press, 1983, Bayesian Inference for Pareto Populations. *Journal of Econometrics* 21, 287-306.

[41] Bajari, Patrick, 1997, Structural Econometric Model of the First-Price, Sealed-Bid Auction: With Applications to Procurement of Highway Improvements. Unpublished manuscript.

[41] Berger, J.O., Liseo, B. and R.L. Wolpert, 1998, Integrated Likelihood Methods for Eliminating Nuisance Parameters. Working Paper 9701, Institute of Statistics and Decision Sciences, Duke University.

[41] Berger, J.O. and J.M. Bernardo, 1992, On the development of Reference Priors. Bayesian Statistics 4 (J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, eds.), Oxford University Press, 35-60 (with discussion).

[41] Bernardo, J.M. and A.F.M. Smith, 1994. *Bayesian Theory*, John Wiley and Sons, Inc., New York.

[41] Bernardo, J.M., 1979, Reference Posterior Distributions for Bayesian Inference. *J. Roy. Statist. Soc. B*, 41, 113-147 (with discussion).

[41] Castillo, E. and Ali S. Hadi, 1997, Fitting the Generalized Pareto Distribution to Data. *Journal of the American Statistical Association* 92, 1609-1620.

[41] Christensen, B.J. and A.N. Kiefer, 1991, The Exact Likelihood Function for an Empirical Job Search Model. *Econometric Theory* 7, 464-486.

[41] Dawid, A.P., 1979, Conditional Independence in Statistical Theory (with discussion). *The Journal of the Royal Statistical Society*, Series B, 41, 1-31.

[41] Stephen G. Donald and H.J. Paarsch, Identification, Estimation, and Testing in Parametric Empirical Models of Auctions within the Independent Private Values Paradigm. *Econometric Theory* 12, 517-567.

[41] Galambos, J.,1978. *The Asymptotic Theory of Extreme Order Statistics*, New York, Wiley.

[41] Ghosal, S., 1997, Reference Priors in Multiparameter Nonregular case. *Test* 6, 159-186.

[41] Ghosal, S. and T. Samanta, 1997, Expansion of Bayes risk for Entropy Loss and Reference Prior in Nonregular Case. *Statistics and Decisions* (to appear).

[41] Ghosal, S. and T. Samanta, 1995, Asymptotic Behavior of Bayes Estimates and Posterior Distributions in Multiparameter Nonregular Case. *Math. Methods Statist.* 4, 361-388.

[41] Ghosh, J.K., Ghosal S. and T. Samanta, 1994, Stability and Convergence of Posterior in Nonregular Problems. *Statistical Decision Theory and Related Topics V* (S.S Gupta and J.O. Berger, eds). New York, Springer-Verlag, 183-199.

[41] Ghosh, M, Ghosh, A., Chen, M.H. and A. Angresti, 1999, Noninformative Priors for One Parameter Item Response Models, unpublished manuscript, Department of Statistics, University of Florida, Gainesville.

[41] Heckman, J. and Honore, B., 1990, The Empirical Content of the Roy Model. *Econometrica* 58, 1121-1149.

[41] Hong, H., 1997, Nonregular Maximum Likelihood Estimation in Auction, Job Search and Production Frontier Models. Unpublished manuscript.

[41] Huzurbazar, V.S. 1976. *Sufficient Statistics: Selected Contributions.* New York, Marcel Dekker.

[41] Ibragimov, I.A. and R.Z. Has'minskii, 1981. *Statistical Estimation: Asymptotic Theory.* New York, Springer-Verlag.

[41] Johnson, N.L., S. Kotz and N. Balakrishnan, 1994. *Continuous Univariate Distributions*, Second edition, Volume 1,2. Wiley, New York.

[41] Kass, R.E. and L. Wasserman, 1996, The Selection of Prior Distributions by Formal Rules. *Journal of the American Statistical Association* 91, 1343-1996.

[41] Koop, G. and D. J. Poirier, 1997, Learning About the Across-Regime Correlation in Switching Regression Models. *Journal of Econometrics* 78, 217-227.

[41] Laffont, J.J., Ossard, H. and Q. Vuong, 1995, Econometrics of First-Price Auctions. *Econometrica* 63, 953-980.

[41] Lancaster, T., 1997, Exact Structural Inference in Optimal Job-Search Models. *Journal of Business and Economic Statistics* 15, 165-179.

[41] Lehmann, E. L., 1983. *Theory of Point Estimation*. New York, Wiley.

[41] Lindley, D.V., 1956, On a Measure of Information Provided by an Experiment. *Annals of Mathematical Statistics* 27, 986-1005.

[41] Madddala, G.S., 1983. *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge University Press.

[41] Murray, J.D., 1974. *Asymptotic Analysis*. Oxford, U.K., Oxford University Press.

[41] Paarsch, H.J., 1992, Deciding Between the Common and Private Value Paradigms in Empirical Models of Auctions. *Journal of Econometrics* 51, 191-215.

[41] Pickands, J., 1975, Statistical Inference Using Extreme Order Statistics, *The Annals of Statistics* 3, 119-131.

[41] Poirier, D.J. and P.A. Rudd, 1981, On the Appropriateness of Endogenous Switching. *Journal of Econometrics* 16, 249-256.

[41] Poirier, D. J., 1995. *Intermediate Statistics and Econometrics: A Comparative Approach*, Cambridge, MIT press.

[41] Porter, R.H. and J.D. Zona, 1993, Detection of Bid Rigging in Procurement Auctions. *Journal of Political Economy* 101, 518-538.

[41] Roy, A.D., 1951, Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers* 3, 135-146.

[41] Rubin, D.B., 1978, Bayesian Inference for Causal Effects. *The Annals of statistics* 6, 34-58.

[41] Sareen, S., 1999a, Posterior Odds Comparison of a Symmetric Low-Price, Sealed-Bid Auction Within the Common Value and the Independent Private Values Paradigm. *Journal of Applied Econometrics* 14, 651-676.

[41] Sareen, S., 1999b, Estimation and Testing of Structural Parametric Sealed-Bid Auctions. Ph.D thesis, University of Toronto.

[41] Tanner, M.A. and W.H. Wong, 1987, The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association* 82, 528-550.

[41] Vijverberg, W.P.M., 1993, Measuring the Unidentified Parameter of the Extended Roy Model of Selectivity. *Journal of Econometrics* 57, 69-89.

APPENDIX

A Proof for Lemma 1

I can write the density function of the q dimensional data vector $\mathbf{V}$ as

$$f_{\mathbf{v}}(\mathbf{v}|\theta) = f_{\mathbf{v}_p^o}(\mathbf{v}_p^o|\theta) f_{\mathbf{v}_{\mathbf{q-p}}^M|\mathbf{v}_{\mathbf{p}}^o}(\mathbf{v}_{q-p}^M|\ \theta). \qquad (A.1)$$

$f_{\mathbf{v}_p^o}(\mathbf{v}_p^o|\theta)$ is the density function of the p dimensional vector of observed data, $\mathbf{V}_p^O$. $f_{\mathbf{v}_{\mathbf{q-p}}^M|\mathbf{v}_{\mathbf{p}}^o}(\mathbf{v}_{q-p}^M|\ \theta)$ is the density function of the $q-p$ dimensional vector of missing data *conditional* on $\mathbf{V}_p^O$. Then I can write the Hessian of the loglikelihood of the q dimensional statistics, $\mathbf{V}$, as

$$\begin{aligned}
\mathbf{H}_{\mathbf{V}}(\theta;\mathbf{V}) &= \frac{\partial^2 \log f_{\mathbf{v}}(\mathbf{v}|\theta)}{\partial\theta\partial\theta'} \\
&= \frac{\partial^2 \log f_{\mathbf{v}_p^o}(\mathbf{v}_p^o|\theta)}{\partial\theta\partial\theta'} + \frac{\partial^2 \log f_{\mathbf{v}_{\mathbf{q-p}}^M|\mathbf{v}_{\mathbf{p}}^o}(\mathbf{v}_{q-p}^M|\ \theta)}{\partial\theta\partial\theta'} \qquad (A.2)
\end{aligned}$$

To obtain $\mathbf{J}_{\mathbf{V}}(\theta)$, I take the expectation of both sides with respect to the density function $f_{\mathbf{v}}(\mathbf{v}|\theta) = f_{\mathbf{v}_p^o}(\mathbf{v}_p^o|\theta) f_{\mathbf{v}_{q-p}^M|\mathbf{v}_p}(\mathbf{v}_{q-p}^M|\ \theta)$, and multiply by -1;

$$-E_{\mathbf{V}|\theta}\left[\mathbf{H}_{\mathbf{V}}(\theta;\mathbf{V})\right] = -E_{\mathbf{V}|\theta}\left[\frac{\partial^2 \log f_{\mathbf{v}_p^o}(\mathbf{v}_p^o|\theta)}{\partial\theta\partial\theta'}\right] - E_{\mathbf{V}|\theta}\left[\frac{\partial^2 \log f_{\mathbf{v}_{\mathbf{q-p}}^M|\mathbf{v}_{\mathbf{p}}^o}(\mathbf{v}_{q-p}^M|\ \theta)}{\partial\theta\partial\theta'}\right]. \qquad (A.3)$$

I now evaluate both terms on the right-hand side of (A.3).

$$\begin{aligned}
E_{\mathbf{V}|\theta}\left[\frac{\partial^2 \log f_{\mathbf{v}_p^o}(\mathbf{v}_p^o|\theta)}{\partial\theta\partial\theta'}\right] &= \int \cdots \int_{\zeta_{\mathbf{V}}} \left[\frac{\partial^2 \log f_{\mathbf{v}_p^o}(\mathbf{v}_p^o|\theta)}{\partial\theta\partial\theta'}\right] f_{\mathbf{v}}(\mathbf{v}|\theta) dv_1^o \ldots dv_p^o dv_{p+1}^M \ldots dv_q^M, \\
&= \int \cdots \int_{\zeta_{\mathbf{V}_p^O}} \left[\frac{\partial^2 \log f_{\mathbf{v}_p^o}(\mathbf{v}_p^o|\theta)}{\partial\theta\partial\theta'}\right] \left\{\int \cdots \int_{\zeta_{\mathbf{V}_{q-p}^M}} f_{\mathbf{v}_{q-p}^M}(\mathbf{v}_{q-p}^M|\theta) dv_{p+1}^M \ldots dv_q^M\right\} f_{\mathbf{v}_p^o}(\mathbf{v}_p^o|\theta) dv_1^o \ldots dv_p^o, \\
&= \int \cdots \int_{\zeta_{\mathbf{V}_p^O}} \left[\frac{\partial^2 \log f_{\mathbf{v}_p^o}(\mathbf{v}_p^o|\theta)}{\partial\theta\partial\theta'}\right] f_{\mathbf{v}_p^o}(\mathbf{v}_p^o|\theta) dv_1^o \ldots dv_p^o, \\
&= \mathbf{J}_{\mathbf{V}_p^O}(\theta). \qquad (A.4)
\end{aligned}$$

The second term on the right-hand side of (A.3) is now evaluated.

$$\begin{aligned}
& E_{\mathbf{V}|\theta}\left[\frac{\partial^2 \log f_{\mathbf{v}_{\mathbf{q-p}}^M|\mathbf{v}_{\mathbf{p}}^o}(\mathbf{v}_{q-p}^M|\ \theta)}{\partial\theta\partial\theta'}\right] \\
&= \int \cdots \int_{\zeta_{\mathbf{V}}} \left[\frac{\partial^2 \log f_{\mathbf{v}_{\mathbf{q-p}}^M|\mathbf{v}_{\mathbf{p}}^o}(\mathbf{v}_{q-p}^M|\ \theta)}{\partial\theta\partial\theta'}\right] f_{\mathbf{v}}(\mathbf{v}|\theta) dv_1^o \ldots dv_p^o dv_{p+1}^M \ldots dv_q^M, \\
&= \int \cdots \int_{\zeta_{\mathbf{V}_p^O}} \left\{\int \cdots \int_{\zeta_{\mathbf{V}_{q-p}^M}} \left[\frac{\partial^2 \log f_{\mathbf{v}_p^o}(\mathbf{v}_p^o|\theta)}{\partial\theta\partial\theta'}\right] f_{\mathbf{v}_{q-p}^M}(\mathbf{v}_{q-p}^M|\theta) dv_{p+1}^M \ldots dv_q^M\right\} f_{\mathbf{v}_p^o}(\mathbf{v}_p^o|\theta) dv_1^o \ldots dv_p^o, \\
&= E_{\mathbf{V}_p^O|\theta}\left[\mathbf{J}_{\mathbf{V}_{q-p}^M|\mathbf{V}_p^O}(\theta)\right]. \qquad (A.5)
\end{aligned}$$

$\zeta_{\mathbf{V}_p^o}$ and $\zeta_{\mathbf{V}_{q-p}^M}$ are the supports of the density functions $f_{\mathbf{v}_p^o}(\mathbf{v}_p^o|\theta)$ and $f_{\mathbf{v}_{q-p}^M}(\mathbf{v}_{q-p}^M|\theta)$, respectively. Substituting (A.4) and (A.5) in (A.3), I get

$$\mathbf{J}_{\mathbf{V}}(\theta) = \mathbf{J}_{\mathbf{V}_p}(\theta) + E_{\mathbf{V}_p|\theta}\left[\mathbf{J}_{\mathbf{V}_{nT-p}|\mathbf{V}_p}(\theta)\right].$$

B Proof for Corollary 1

I prove Corollary 1 in this Appendix. I first prove

$$\mathbf{V}_p^O \quad \text{jointly sufficient for} \quad \theta \Rightarrow \mathbf{J}_{\mathbf{V}}(\theta) = \mathbf{J}_{\mathbf{V}_p^O}(\theta).$$

The score function of the density function of $\mathbf{V}$ given in (A.1) is

$$\begin{aligned}
\mathbf{S}_{\mathbf{V}}(\theta;\mathbf{V}) &= \frac{\partial \log f_{\mathbf{v}}(\mathbf{v}|\theta)}{\partial\theta} = \frac{\partial \log f_{\mathbf{v}_p^o}(\mathbf{v}_p|\theta)}{\partial\theta} + \frac{\partial \log f_{\mathbf{v}_{q}^M|\mathbf{v}_p}(\mathbf{v}_{q-p}^M|\theta)}{\partial\theta} \\
&= \frac{\partial \log f_{\mathbf{v}_p^o}(\mathbf{v}_p|\theta)}{\partial\theta} = \mathbf{S}_{\mathbf{V}_p^O}\left(\theta;\mathbf{V}_p^O\right).
\end{aligned}$$

Since $\mathbf{V}_p^O$ are jointly sufficient for θ, by definition of sufficiency (Huzurbazar, 1976, pp. 105-108), the distribution of arbitrary statistics, $\mathbf{V}_{q-p}^M$ conditional on the jointly sufficient statistics, $\mathbf{V}_p^O$, will not depend on θ; hence $\frac{\partial \log f_{\mathbf{v}_{n_j-p}|\mathbf{v}_p}(\mathbf{v}_{n_j-p}|\theta)}{\partial\theta}$ will be zero. Then

$$\begin{aligned}
\mathbf{J}_{\mathbf{V}}(\theta) &= \int ... \int_{\zeta_{\mathbf{V}}} \left[\mathbf{S}_{\mathbf{V}}(\theta;\mathbf{V})\right]\left[\mathbf{S}_{\mathbf{V}}(\theta;\mathbf{V})\right]' f_{\mathbf{v}}(\mathbf{v}|\theta) dv_1^o...dv_p^o dv_{p+1}^M...dv_q^M, \\
&= \int ... \int_{\zeta_{\mathbf{V}_p^O}} \left[\mathbf{S}_{\mathbf{V}_p^O}\left(\theta;\mathbf{V}_p^O\right)\right]\left[\mathbf{S}_{\mathbf{V}_p^O}\left(\theta;\mathbf{V}_p^O\right)\right]' f_{\mathbf{v}_p^o}(\mathbf{v}_p^o|\theta) dv_1^o...dv_p^o, \\
&= \mathbf{J}_{\mathbf{V}_p^O}(\theta).
\end{aligned}$$

Now I prove

$$\mathbf{J}_{\mathbf{V}}(\theta) = \mathbf{J}_{\mathbf{V}_p^O}(\theta) \Rightarrow \mathbf{V}_p^O \quad \text{jointly sufficient for} \quad \theta.$$

From Lemma 1

$$\mathbf{J}_{\mathbf{V}}(\theta) = \mathbf{J}_{\mathbf{V}_p^O}(\theta) + E_{\mathbf{V}_p^O|\theta}\left[\mathbf{J}_{\mathbf{V}_{q-p}^M|\mathbf{V}_p^O}(\theta)\right];$$

hence $\mathbf{J}_{\mathbf{V}}(\theta) = \mathbf{J}_{\mathbf{V}_p^O}(\theta)$ implies that $E_{\mathbf{V}_p^O|\theta}\left[\mathbf{J}_{\mathbf{V}_{q-p}^M|\mathbf{V}_p^O}(\theta)\right]$ is a null matrix. Now

$$E_{\mathbf{V}_p^O|\theta}\left[\mathbf{J}_{\mathbf{V}_{q-p}^M|\mathbf{V}_p^O}(\theta)\right] = E_{\mathbf{V}_p^O|\theta}\left[\mathbf{S}_{\mathbf{V}_{q-p}^M|\mathbf{V}_p^O}\left(\theta;\mathbf{V}_{q-p}^M\right)\mathbf{S}'_{\mathbf{V}_{q-p}^M|\mathbf{V}_p^O}\left(\theta;\mathbf{V}_{q-p}^M\right)\right],$$

where

$$\mathbf{S}_{\mathbf{V}_{q-p}^M|\mathbf{V}_p^O}\left(\theta;\mathbf{V}_{q-p}^M\right) = \left[\frac{\partial \log f_{\mathbf{V}_{q-p}^M|\mathbf{V}_p^O}(\mathbf{v}_{q-p}^M|\theta)}{\partial\theta_1}.....\frac{\partial \log f_{\mathbf{V}_{q-p}^M|\mathbf{V}_p^O}(\mathbf{v}_{q-p}^M|\theta)}{\partial\theta_k}\right]'$$

is a k dimensional vector of score functions of $f_{\mathbf{V}_{q-p}^M|\mathbf{V}_p^O}(\mathbf{v}_{q-p}^M|\theta)$. A typical diagonal element of $E_{\mathbf{V}_p^O|\theta}\left[\mathbf{J}_{\mathbf{V}_{q-p}^M|\mathbf{V}_p^O}(\theta)\right]$ is $E_{\mathbf{V}_p^O|\theta}\left[\left(\frac{\partial \log f_{\mathbf{V}_{q-p}^M|\mathbf{V}_p^O}(\mathbf{v}_{q-p}^M|\theta)}{\partial\theta_\tau}\right)^2\right]$, $\tau = 1,...,k$; for these diagonal elements to be zero, a necessary and sufficient condition is that the distribution of $\mathbf{V}_{q-p}^M$ conditional on $\mathbf{V}_p^O$ is independent of $\theta_\tau, \forall\ \tau = 1,...,k$; but then $\mathbf{V}_p^O$ are jointly sufficient statistics for θ by definition of a sufficient statistic (Huzurbazar, 1976, pp. 105-108).

C Fisher Information for Structural Auction Model

When bids are proportional to cost as in Section 4.2, the Fisher information matrix for (η,φ) from the bids and the winning bid are given in equation (C.1) and (C.2), respectively,

$$\mathbf{J_B}(\eta,\varphi) = \begin{bmatrix} (nT)^2\frac{\varphi^2}{\eta^2} & -\frac{(nT)^2\varphi}{\eta(\varphi m-1)} \\ -\frac{(nT)^2\varphi}{\eta(\varphi m-1)} & \frac{nT(\varphi m-1)^2+(nT)^2\varphi^2}{\varphi^2(\varphi m-1)^2} \end{bmatrix}, \qquad (C.1)$$

$$\text{and} \quad \mathbf{J_W}(\eta,\varphi) = \begin{bmatrix} (nT)^2\frac{\varphi^2}{\eta^2} & -\frac{(nT)^2\varphi}{\eta(\varphi m-1)} \\ -\frac{(nT)^2\varphi}{\eta(\varphi m-1)} & \frac{T(\varphi m-1)^2+(nT)^2\varphi^2}{\varphi^2(\varphi m-1)^2} \end{bmatrix}, \quad C.2$$

where $m = n-1$. Hence

$$\mathbf{J_B}(\eta,\varphi) - \mathbf{J}_W(\eta,\varphi) = \begin{bmatrix} 0 & 0 \\ 0 & \frac{T(n-1)}{\varphi} \end{bmatrix}, \qquad (C.3)$$

which is *psd*.

The Fisher information for (η,φ) from the minimum winning bid across auctions is

$$\mathbf{J}_{W_*}(\eta,\varphi) = \begin{bmatrix} (nT)^2\frac{\varphi^2}{\eta^2} & -\frac{(nT)^2\varphi}{\eta(\varphi m-1)} \\ -\frac{(nT)^2\varphi}{\eta(\varphi m-1)} & \frac{(\varphi m-1)^2+(nT)^2\varphi^2}{\varphi^2(\varphi m-1)^2} \end{bmatrix}.$$

D Fisher Information for Job-Search Model

I first obtain $E_{\mathbf{W},\mathbf{t}|\theta}\left[\mathbf{J}^{\xi\xi}_{\mathbf{U},\mathbf{S}|\mathbf{W},\mathbf{t}}(\theta)\right]$. The density function of the missing data conditional on the observed data follows from Lancaster (1997, p. 166),

$$f_{us|wt}(\mathbf{u}_j, s_j|w_j,t_j) = \prod_{i=1}^{s_j}(\lambda t_j)^{s_j}e^{-\lambda F_w(\xi|\mu,\sigma)t_j}f_w(u_j^i|\mu,\sigma)/s_j!, \quad u_j^1,...,u_j^{s_j} < \xi.$$

The likelihood function across the N *iid* agents is

$$L_{\mathbf{U},\mathbf{S}|\mathbf{W},\mathbf{t}}(\theta;\mathbf{U},\mathbf{S}) = \left[\prod_{j=1}^{N}\prod_{i=1}^{s_j}f_w(u_j^i|\mu,\sigma)\right]\left[\prod_{j=1}^{N}(\lambda t_j)^{s_j}\right]\left[e^{-\lambda TF_w(\xi|\mu,\sigma)}\right]/\prod_{j=1}^{N}s_j!.$$

As a result the score function is

$$S_{\mathbf{U},\mathbf{S}|\mathbf{W},\mathbf{t}}(\xi;\mathbf{U},\mathbf{S}) = \frac{\partial f_{us|wt}(\bullet)}{\partial \xi} = (\lambda T) f_w(\xi|\mu,\sigma).$$

Then

$$\begin{aligned}\mathbf{J}^{\xi\xi}_{\mathbf{U},\mathbf{S}|\mathbf{W},\mathbf{t}}(\theta) &= \mathbf{E}_{\mathbf{U},\mathbf{S}|\mathbf{W},\mathbf{t}}\left[(\lambda T f_w(\xi|\mu,\sigma))^2\right], \\ &= (\lambda T f_w(\xi|\mu,\sigma))^2 .\end{aligned}$$

Hence

$$E_{\mathbf{W},\mathbf{t}|\theta}\left[\mathbf{J}^{\xi\xi}_{\mathbf{U},\mathbf{S}|\mathbf{W},\mathbf{t}}(\theta)\right] = \lambda^2 \left[f_w(\xi|\mu,\sigma)\right]^2 E_{\mathbf{W},\mathbf{t}|\theta}\left[T^2\right], \qquad (D.1)$$

where $T = \sum_{j=1}^{N} t_j$. I will now evaluate $E_{\mathbf{W},\mathbf{t}|\theta}\left[T^2\right]$ in (D.1). I will suppress below the conditioning in the density functions on the parameter vector θ.

$$\begin{aligned}E_{\mathbf{W},\mathbf{t}|\theta}\left[T^2\right] &= \int_{w_1} ... \int_{w_N} \int_{t_1} ... \int_{t_N} T^2 f_{wt}(\mathbf{w},\mathbf{t}) dw_1 ... dw_N dt_1 ... dt_N, \\ &= \int_{t_1} ... \int_{t_N} T^2 \left[\int_{w_1} ... \int_{w_N} f_{\mathbf{w}|\mathbf{t}}(\mathbf{w}|\mathbf{t}) dw_1 ... dw_N\right] f_{\mathbf{t}}(\mathbf{t}) dt_1 ... dt_N, \\ &= \int_{t_1} ... \int_{t_N} T^2 f_{\mathbf{t}}(\mathbf{t}) dt_1 ... dt_N, \\ &= \int_{t_1} ... \int_{t_N} \left[\sum_{j=1}^{N} t_j^2 + 2 \sum_{\tau,\tau'} t_\tau t_{\tau'}\right] f_{\mathbf{t}}(\mathbf{t}) dt_1 ... dt_N, \qquad (D.2)\end{aligned}$$

where $\tau \neq \tau'$, $\tau, \tau' = 1, ..., N$. Since the t_j, $j = 1, ..., N$ are independently and identically distributed, (D.2) simplifies to

$$E_{\mathbf{W},\mathbf{t}|\theta}\left[T^2\right] = N \int_{t_j} t_j^2 f_t(t_j) dt_j + N(N-1) \left[\int_{t_j} t_j f_t(t_j) dt_j\right]^2 . \qquad (D.3)$$

Exploiting the fact (Lancaster, 1997, p. 166) that $t_j \sim EXP(\lambda \; \bar{F}_w \; (\xi|\mu,\sigma))$,

$$E_t(t_j) = \lambda \; \bar{F}_w \; (\xi|\mu,\sigma), \quad E_t(t_j^2) = 2 \left[\lambda \; \bar{F}_w \; (\xi|\mu,\sigma)\right]^2 .$$

I substitute these in (D.3), and the resulting expression in (D.1) to obtain

$$E_{\mathbf{W},\mathbf{t}|\theta}\left[\mathbf{J}^{\xi\xi}_{\mathbf{U},\mathbf{S}|\mathbf{W},\mathbf{t}}(\theta)\right] = N(N+1) \left[\lambda f_w(\xi|\mu,\sigma)\right]^2 \left[\lambda \; \bar{F}_w \; (\xi|\mu,\sigma)\right]^2 .$$

Clearly this is nonzero; hence $\mathbf{J}^{\xi\xi}_{\mathbf{U},\mathbf{S},\mathbf{W},\mathbf{t}}(\theta) \neq \mathbf{J}^{\xi\xi}_{\mathbf{W},\mathbf{t}}(\theta)$.

E Reference Prior Under Job-Search Model

The sampling expectation of the score of the density function under *full observability* is

$$c_{suwt}(\xi) = E_{S\mathbf{U}WT|\theta}\left(\frac{\partial \log f_{s\mathbf{u}wt}(s_j, \mathbf{u}_j, w_j, t_j|\xi, \varphi)}{\partial \eta}\right),$$

where $f_{s\mathbf{u}wt}(s_j, \mathbf{u}_j, w_j, t_j|\xi, \varphi)$ is given by equation (4). Since $U_j^i|S_j, W_j, T_j$ are independently and identically distributed for an agent j, and reference priors are defined upto a constant of proportionality, the *proposed* noninformative prior for ξ given φ under *full observability* is proportional to

$$\pi^{suwt}(\xi|\varphi) \propto E_{SUWT|\theta}\left(\frac{\partial \log f_{suwt}(s_j, u_j^i, w_j, t_j|\xi, \varphi)}{\partial \xi}\right). \qquad (E.1)$$

This is now obtained. I will indicate $\theta = (\xi, \varphi)$.

$$\frac{\partial \log f_{suwt}(\bullet|\theta)}{\partial \xi} = \frac{\partial}{\partial \xi} \log\left[f_t(t_j|\theta) f_{w|t}(w_j|t_j, \theta) f_{s|wt}(s_j|w_j, t_j, \theta) f_{u|swt}(s_j|w_j, t_j, \theta)\right].$$

Taking expectation on both sides with respect to the sampling density $SUWT|\theta$, I obtain

$$E_{SUWT|\theta}\left(\frac{\partial \log f_{suwt}(s_j, u_j^i, w_j, t_j|\theta)}{\xi}\right) = E_{SUWT|\theta}\left[\frac{\partial \log f_t(t_j|\theta)}{\partial \xi}\right] +$$

$$E_{SUWT|\theta}\left[\frac{\partial \log f_{w|t}(w_j|t_j, \theta)}{\partial \xi}\right] + E_{SUWT|\theta}\left[\frac{\partial \log f_{s|wt}(s_j|w_j, t_j, \theta)}{\partial \xi}\right]$$
$$+E_{SUWT|\theta}\left[\frac{\partial \log f_{u|swt}(u_j^i|s_j, w_j, t_j, \theta)}{\partial \xi}\right] \qquad (E.2)$$

Each term in (E.2) is now obtained.

$$\begin{aligned}
E_{SUWT|\theta}\left[\frac{\partial \log f_t(t_j|\theta)}{\partial \xi}\right] &= \int_{t_j}\int_{w_j}\int_{u_j^i}\sum_{s_j}\frac{\partial \log f_t(t_j|\theta)}{\partial \xi} f_{suwt}(s_j, u_j^i, w_j, t_j|\theta)\, du_j^i\, dw_j\, dt_j, \\
&= \int_{t_j}\int_{w_j}\int_{u_j^i}\sum_{s_j}\frac{\partial f_t(t_j|\theta)}{\partial \xi} f_{suw|t}(s_j, u_j^i, w_j|t_j, \theta)\, ds_j\, du_j^i\, dw_j\, dt_j, \\
&= \int_{t_j}\left[\int_{w_j}\int_{u_j^i}\sum_{s_j} f_{suw|t}(s_j, u_j^i, w_j|t_j, \theta)\, du_j^i\, dw_j\right]\frac{\partial f_t(t_j|\theta)}{\partial \xi} dt_j, \\
&= \int_{t_j}\frac{\partial f_t(t_j|\theta)}{\partial \xi} dt_j \\
&= \frac{\partial}{\partial \xi}\int_{t_j} f_t(t_j|\theta)\, dt_j = 0. \qquad (E.3)
\end{aligned}$$

Since the support of the density function of T_j does not depend on any parameter, the interchange of integration and differentiation is valid. The second term is now evaluated.

$$E_{SUWT|\theta}\left[\frac{\partial \log f_{w|t}(w_j|t_j,\theta)}{\partial \xi}\right] = \int_{t_j}\int_{w_j}\int_{u_j^i}\sum_{s_j}\frac{\partial \log f_{w|t}(w_j|t_j,\theta)}{\partial \xi} f_{suwt}(s_j,u_j^i,w_j,t_j|\theta)du_j^i dw_j dt_j,$$

$$= \int_{t_j}\int_{w_j}\int_{u_j^i}\sum_{s_j}\frac{\partial f_{w|t}(w_j|t_j,\theta)}{\partial \xi} f_t(t_j|\theta) f_{su|wt}(s_j,u_j^i|w_j,t_j,\theta)du_j^i dw_j dt_j,$$

$$= \int_{t_j}\int_{w_j}\left[\int_{u_j^i}\sum_{s_j} f_{su|wt}(s_j,u_j^i|w_j,t_j,\theta)du_j^i\right]\frac{\partial f_{w|t}(w_j|t_j,\theta)}{\partial \xi} f_t(t_j|\theta)dw_j dt_j,$$

$$= \int_{t_j}\left[\int_{w_j=\xi}^{\infty}\frac{\partial f_{w|t}(w_j|t_j,\theta)}{\partial \xi}dw_j\right] f_t(t_j|\theta)dt_j;$$

applying Leibnitz's rule to evaluate the integral in the square brackets,

$$= \frac{fw(\xi|\varphi)}{\bar{F_w}\ (\xi|\varphi)}\int_{t_j} f_t(t_j|\theta)dt_j = \frac{fw(\xi|\varphi)}{\bar{F_w}\ (\xi|\varphi)}, \qquad (E.4)$$

where $f_{w|t}(w_j|t_j,\theta)$ has been obtained from Lancaster (1997, p. 166). The interchange of integrals and of the integrals and the summation sign is valid from Fubini's theorem. The third term is now evaluated.

$$E_{SUWT|\theta}\left[\frac{\partial \log f_{s|wt}(s_j|w_j,t_j,\theta)}{\partial \xi}\right]$$

$$= \int_{t_j}\int_{w_j}\int_{u_j^i}\sum_{s_j}\frac{\partial \log f_{s|wt}(s_j|w_j,t_j,\theta)}{\partial \xi} f_{suwt}(s_j,u_j^i,w_j,t_j|\theta)du_j^i dw_j dt_j,$$

$$= \int_{t_j}\int_{w_j}\int_{u_j^i}\left[\sum_{s_j}\frac{\partial f_{s|wt}(s_j|w_j,t_j,\theta)}{\partial \xi}\right] f_{wt}(w_j,t_j|\theta) f_{u|swt}(u_j^i|s_j,w_j,t_j,\theta)du_j^i dw_j dt_j,$$

$$= \int_{t_j}\int_{w_j}\int_{u_j^i}\left[\frac{\partial}{\partial \xi}\sum_{s_j}\frac{f_{s|wt}(s_j|w_j,t_j,\theta)}{\partial \xi}\right] f_{wt}(w_j,t_j|\theta) f_{u|swt}(u_j^i|s_j,w_j,t_j,\theta)du_j^i dw_j dt_j,$$

$$= 0. \qquad (E.5)$$

The last term in (E.2) now evaluated.

$$E_{SUWT|\theta}\left[\frac{\partial \log f_{u|swt}(u_j^i|s_j,w_j,t_j,\theta)}{\partial \xi}\right]$$

$$= \int_{t_j}\int_{w_j}\int_{u_j^i}\sum_{s_j}\frac{\partial \log f_{u|swt}(u_j^i|s_j,w_j,t_j,\theta)}{\partial \xi} f_{suwt}(s_j,u_j^i,w_j,t_j|\theta)du_j^i dw_j dt_j,$$

$$= \int_{t_j}\int_{w_j}\int_{u_j^i}\sum_{s_j}\frac{\partial f_{u|swt}(u_j^i|s_j,w_j,t_j,\theta)}{\partial \xi} f_{swt}(s_j,w_j,t_j|\theta)du_j^i dw_j dt_j,$$

$$= \int_{t_j}\int_{u_j^i}\sum_{s_j}\left[\int_{u_j^i=0}^{w_j}\frac{\partial f_{u|swt}(u_j^i|s_j,w_j,t_j,\theta)}{\partial\xi}du_j^i\right]f_{swt}(s_j,w_j,t_j|\theta)dw_jdt_j;$$

applying Leibnitz's rule to evaluate the integral in the square brackets,

$$= -\frac{fw(\xi|\varphi)}{F_w(\xi|\varphi)}, \qquad (E.6)$$

where $f_{u|swt}(u_j^i|s_j,w_j,t_j,\theta)$ has been obtained from Lancaster (1997, p. 166). Substituting (E.3)-(E.6) in (E.2) obtains the *proposed* noninformative prior for ξ given φ under *full observability* in equation (33).

Under *partial observability*, the proposed noninformative prior for ξ given φ is

$$\pi^{wt}(\xi|\varphi) \propto E_{WT|\theta}\left(\frac{\partial\log f_{wt}(w_j,t_j|\theta)}{\partial\xi}\right).$$

The density function of the observed data, $f_{wt}(w_j,t_j|\theta)$, is given in equation (5). The score of this density function is

$$\frac{\partial\log f_{wt}(w_j,t_j|\theta)}{\partial\xi} = \frac{\partial}{\partial\xi}\log\left[f_t(t_j|\theta)f_{w|t}(w_j|t_j,\theta)\right];$$

and the sampling expectation of this score is

$$E_{WT|\theta}\left(\frac{\partial\log f_{wt}(w_j,t_j|\theta)}{\partial\xi}\right) = E_{WT|\theta}\left(\frac{\partial\log f_t(t_j|\theta)}{\partial\xi}\right) + E_{WT|\theta}\left(\frac{\partial\log f_{w|t}(w_j|t_j,\theta)}{\partial\xi}\right).$$

Evaluating each sampling expectation on the right-hand side in a manner analogous to the proposed noninformative prior under *full observability*, I obtain the expression for the *proposed* noninformative prior for ξ given φ in equation (38).

F Fisher Information for Roy Selection Model

The log likelihood function for the Roy model is

$$\log f_{I^*,Y_1^*,Y_2^*}(\mathbf{z}_j|\theta) = -\frac{1}{2}\log|\mathbf{\Omega}| - \frac{1}{2}\mathbf{z}_j'\mathbf{\Omega}^{-1}\mathbf{z}_j, \qquad (F.1)$$

where $\mathbf{z}_j = [z_{Ij}, z_{1j}, z_{2j}]' = [I_j^* - \mathbf{W}_j\gamma, Y_{1j}^* - \mathbf{X}_{1j}\beta_1, Y_{2j}^* - \mathbf{X}_{2j}\beta_2]'$. $|\mathbf{\Omega}|$ is the determinant of the covariance matrix $\mathbf{\Omega}$; it is restricted to be *strictly* greater than zero from the positive definiteness of $\mathbf{\Omega}$. I will indicate it by D,

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

Simplifying $\mathbf{z}_j'\mathbf{\Omega}^{-1}\mathbf{z}_j$ I get

$$\mathbf{z}_j'\mathbf{\Omega}^{-1}\mathbf{z}_j = \frac{A}{D}, \qquad (F.3)$$

where

$$\begin{aligned} A &= (1-\rho_{12}^2)z_{Ij}^2 + (1-\rho_{2u}^2)z_{1j}^2 + (1-\rho_{1u}^2)z_{2j}^2 + \\ & 2(\rho_{12}\rho_{2u}-\rho_{1u})z_{Ij}z_{1j} + 2(\rho_{12}\rho_{1u}-\rho_{2u})z_{Ij}z_{2j} + \\ & 2(\rho_{1u}\rho_{2u}-\rho_{12})z_{1j}z_{2j}. \qquad (F.4) \end{aligned}$$

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

$$\log f_{I^*,Y_1^*,Y_2^*}(\mathbf{z}_j|\theta) = -\frac{1}{2}\log D - \frac{1}{2}\frac{A}{D}. \qquad (F.5)$$

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

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

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

$$\begin{aligned} H^{\rho_{12}\rho_{12}}_{I^*,Y_1^*,Y_2^*}(\theta) &= -\frac{1}{2D}\left[\frac{d^2D}{d\rho_{12}^2}\left(1-\frac{A}{D}\right) + \frac{d^2A}{d\rho_{12}^2} + \frac{1}{D}\left(\frac{dD}{d\rho_{12}}\right)^2\left(\frac{2A}{D}-1\right)\right. \\ & \left. -\frac{2}{D}\frac{dA}{d\rho_{12}}\frac{dD}{d\rho_{12}}\right]. \end{aligned}$$

Obtaining the derivatives in the above expression from (F.2) and (F.4), and taking the expectation of each of these terms with respect to the sampling density $I^*, Y_1^*, Y_2^*|\theta$ I obtain the Fisher information about ρ_{12} under *full observability*, $J^{\rho_{12}\rho_{12}}_{\mathbf{I}^*,\mathbf{Y}_1^*,\mathbf{Y}_2^*}(\theta)$, given by (20).

G Reference Prior for Roy Selection Model

In this Appendix I prove that the reference posterior given in (43) is integrable. From equations (42) and (44)

$$\int \pi^{I^*Y_1^*Y_2^*}(\rho_{12}|\rho_{1u},\rho_{2u},\mathbf{Y}_1^O,\mathbf{Y}_2^O,\mathbf{I})d\rho_{12} \propto \int_{\rho_{12}^l}^{\rho_{12}^u}\sqrt{\left[\frac{1}{D}+\frac{2}{D^2}\left(\rho_{1u}\rho_{2u}-\rho_{12}\right)^2\right]}d\rho_{12}, \qquad (G.1)$$

where

$$\rho_{12}^l = \rho_{1u}\rho_{2u} - \sqrt{(1-\rho_{1u}^2)(1-\rho_{2u}^2)}, \quad \rho_{12}^u = \rho_{1u}\rho_{2u} + \sqrt{(1-\rho_{1u}^2)(1-\rho_{2u}^2)}. \qquad (G.2)$$

Now

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

I now show that each term on the right-hand side is bounded.

Starting with the first integral, since $D \to 0$, as $\rho_{12} \to \rho_{12}^l, \rho_{12}^u$,

$$\int_{\rho_{12}^l}^{\rho_{12}^u} \sqrt{\frac{1}{D}} d\rho_{12} = \lim_{\epsilon^l \to \rho_{12}^l} \lim_{\epsilon^u \to \rho_{12}^u} \int_{\epsilon^l}^{\epsilon^u} \sqrt{\frac{1}{D}} d\rho_{12}. \qquad (G.4)$$

From Gradshteyn, et al. (1993),

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

Then

$$\begin{aligned}
\lim_{\epsilon^l \to \rho_{12}^l} \lim_{\epsilon^u \to \rho_{12}^u} \int_{\epsilon^l}^{\epsilon^u} \sqrt{\frac{1}{D}} d\rho_{12} &= -\arcsin\left(\frac{(\rho_{1u}\rho_{2u} - \rho_{12}^u)}{\sqrt{(1-\rho_{1u}^2)(1-\rho_{2u}^2)}}\right) \\
&\quad + \arcsin\left(\frac{(\rho_{1u}\rho_{2u} - \rho_{12}^l)}{\sqrt{(1-\rho_{1u}^2)(1-\rho_{2u}^2)}}\right), \\
&= 3.14. \qquad (G.5)
\end{aligned}$$

The last step follows from substituting ρ_{12}^l, ρ_{12}^u from (G.2).

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

$$\begin{aligned}
\sqrt{2}\int_{\rho_{12}^l}^{\rho_{12}^u} \frac{(\rho_{1u}\rho_{2u} - \rho_{12})}{D} d\rho_{12} &= \frac{1}{\sqrt{2}} \log D|_{\rho_{12}^l}^{\rho_{12}^u} \\
&< \log D|_{\rho_{12}^l - \epsilon^l}^{\rho_{12}^u + \epsilon^u}, \quad \epsilon^l, \epsilon^u > 0, \\
&= \log \frac{\epsilon^u}{\epsilon^l} \\
&< \infty.
\end{aligned}$$

Since both terms in (G.3) are bounded, the reference posterior $\pi^{I^* Y_1^* Y_2^*}(\rho_{12}|\rho_{1u}, \rho_{2u}, \mathbf{Y}_1^O, \mathbf{Y}_2^O, \mathbf{I})$ is integrable.

Table 1: Definition of Quantities

n = number of participants in an auction.

T = number of auctions in the sample.

$\mathbf{B} = [\mathbf{B}_1, ..., \mathbf{B}_j, ..., \mathbf{B}_T] = n \times T$ dimensional matrix of bids.

$\mathbf{B}_j = [B_j^1, .., B_j^i, ..., B_j^n]' = n$ dimensional vector of bids in auction j.

$W_j = \min\limits_{i=1,...,n} B_j^i$ = winning bid in auction j.

$W_* = \min\limits_{j=1,...,T} W_j$ = minimum winning bid across the T auctions.

$\mathbf{B}_{n:n} = [\mathbf{B}_{1,n:n}, ..., \mathbf{B}_{T,n:n}] = n \times T$ matrix of n ordered bids in T auctions.

$\mathbf{B}_{j,n:n} = \left[B_j^{1:n}, ..., B_j^{i:n}, ..., B_j^{n-1:n}, W_j\right]'$,

$= n$ dimensional vector of ordered bids in auction j, $B_j^{1:n} \leq ... \leq B_j^{n-1:n} \leq W_j$.

$\mathbf{B}_{n-1:n} = [\mathbf{B}_{1,n-1:n}, ..., \mathbf{B}_{j,n-1:n}, ..., \mathbf{B}_{T,n-1:n}]$,

$= n - 1 \times T$ dimensional matrix of $n-1$ ordered bids in T auctions.

$\mathbf{B}_{j,n-1:n} = \left[B_j^{1:n}, ..., B_j^{i:n}, ..., B_j^{n-1:n}\right]'$,

$= n - 1$ dimensional vector of ordered bids in auction j, $B_j^{1:n} \leq ... \leq B_j^{n-1:n}$.

$f_{\mathbf{z}}(\mathbf{z}|\theta)$ = density function of $\mathbf{Z}$ with support $\zeta_{\mathbf{Z}}$.

$\zeta_{\mathbf{Z}(\boldsymbol{\theta})}$ = support of $f_{\mathbf{z}}(\mathbf{z}|\theta)$ if it depends on $\boldsymbol{\theta}$.

$F_z(z|\theta) = \int_{-\infty}^{z} f(\tau|\theta)d\tau$ = distribution function of Z.

$f_{n:n}\left(\mathbf{b}_{j,n:n} \mid \boldsymbol{\theta}\right) = n! \prod_{i=1}^{n} f\left(b_j^{i:n} \mid \boldsymbol{\theta}\right), -\infty < b_j^{1:n} \leq ... \leq w_j < \infty;$

$= \text{joint density of } n \text{ ordered bids, } \mathbf{B}_{n:n}, \text{ in auction } j.$

$f_{n-1:n}\left(\mathbf{b}_{j,n-1:n} \mid \boldsymbol{\theta}\right) = \frac{(n-1)! \prod_{i=1}^{n-1} f\left(b_j^{i::n} \mid \boldsymbol{\theta}\right)}{[F(w_j|\boldsymbol{\theta})]}, -\infty < b_j^{1:n} \leq ... \leq b_j^{n-1:n} \leq w_j;$

$= \text{joint density of } n-1 \text{ ordered bids, } \mathbf{B}_{n-1:n}, \text{ in auction } j.$

$f_w(w_j|\boldsymbol{\theta}) = nf_b(w_j|\boldsymbol{\theta})[1 - F_b(w_j|\boldsymbol{\theta})]^{n-1},$

$= \text{density function of the lowest order statistics from } f_b(b_j^i|\boldsymbol{\theta}).$

$V_i = \nu_i\left(\mathbf{B}_1, .., \mathbf{B}_j, ..., \mathbf{B}_T\right) = \text{an arbitrary statistic.}$

$\mathbf{V} = [V_1, ..., V_{nT}]' = nT \text{ jointly sufficient statistics for } \boldsymbol{\theta}.$

$\ell(\boldsymbol{\theta}; \mathbf{Z}_j) = \sum_{i=1}^{T} \log f(z|\boldsymbol{\theta}) = \text{log-likelihood of } T \text{ observations, } \mathbf{Z}_j.$

$S_Z(\boldsymbol{\theta}; Z) = \frac{\partial \ell(\boldsymbol{\theta}; Z)}{\partial \boldsymbol{\theta}} = \frac{1}{f_z(z|\boldsymbol{\theta})} \frac{\partial f_z(z|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \text{score function of } \ell(\boldsymbol{\theta}; Z).$

$S_{\mathbf{Z}}(\boldsymbol{\theta}; \mathbf{Z}) = \frac{\partial \ell(\boldsymbol{\theta}; \mathbf{Z})}{\partial \boldsymbol{\theta}} = \sum_{j=1}^{T} \sum_{i=1}^{n} S_Z(\boldsymbol{\theta}; Z),$

$= \text{score function of } \ell(\boldsymbol{\theta}; \mathbf{Z}), \mathbf{Z} \text{ being } nT \text{ dimensional.}$

$\mathbf{J}_Z\left(\boldsymbol{\theta}\right) = E_{Z|\boldsymbol{\theta}}\left[\mathbf{S}_Z(\boldsymbol{\theta}; Z)\mathbf{S}_Z'(\boldsymbol{\theta}; Z)\right] = \text{Fisher information from } Z.$

$\mathbf{J}_{\mathbf{Z}}\left(\boldsymbol{\theta}\right) = E_{\mathbf{Z}|\boldsymbol{\theta}}\left[\mathbf{S}_{\mathbf{Z}}\left(\boldsymbol{\theta}; \mathbf{Z}\right) \mathbf{S}_{\mathbf{Z}}'\left(\boldsymbol{\theta}; \mathbf{Z}\right)\right],$

$= \text{Fisher information from } nT \text{ dimensional vector } \mathbf{Z},$

$= nT\mathbf{J}_Z\left(\boldsymbol{\theta}\right) \text{ under standard regularity conditions.}$

$\mathbf{H}_{\mathbf{Z}}(\boldsymbol{\theta}; \mathbf{Z}) = \frac{\partial^2 \ell(\boldsymbol{\theta}; \mathbf{Z})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \text{Hessian matrix of } \mathbf{Z}.$

$m(\bullet), q(\bullet) = \text{strictly positive functions.}$

Figure 1: Reference Prior Kernel for φ

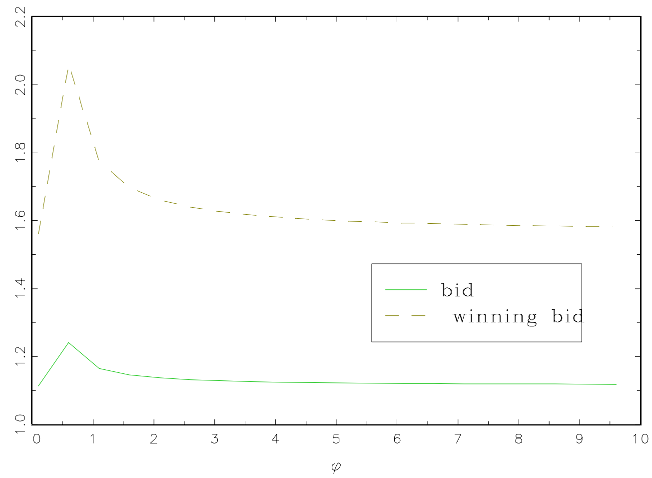

Figure 2: Normalized Reference Prior Density for φ, Truncation at $\varphi = 20$

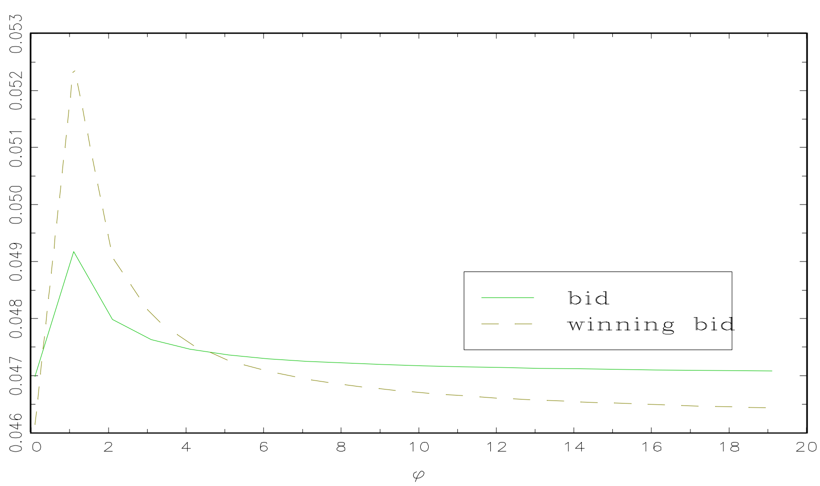

Figure 3: Normalized Reference Prior Density for φ, Truncation at $\varphi = 100$

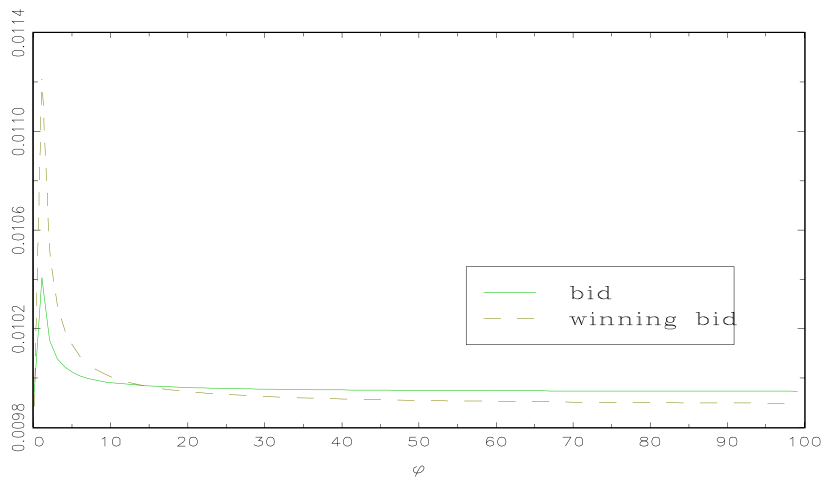

Figure 4: Reference Posterior Kernel from Bids for φ

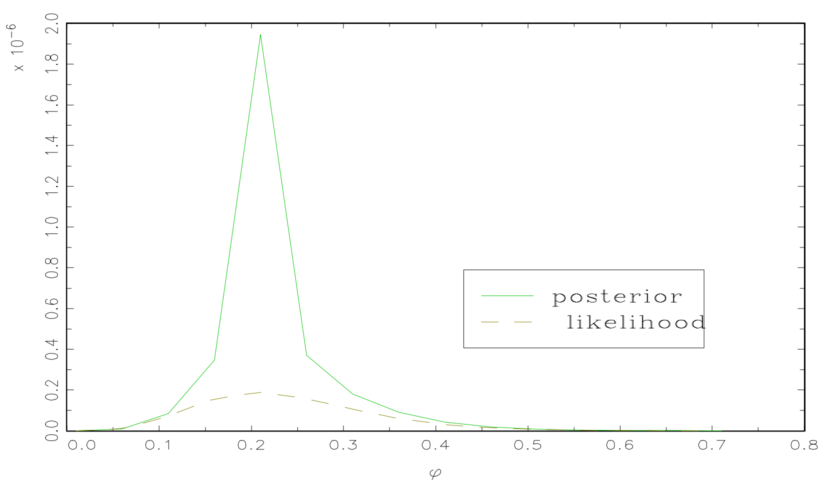

Figure 5: Reference Posterior Kernel from Winning Bids for φ

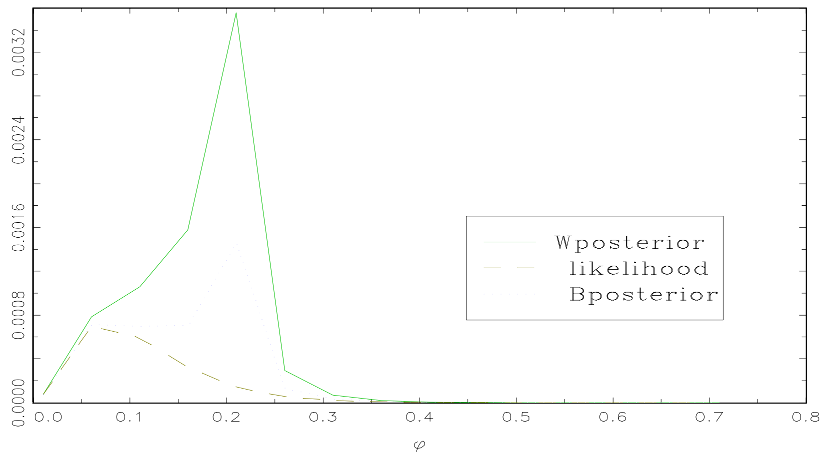

Figure 6: Reference Prior Kernel for ξ in Job-Search Model

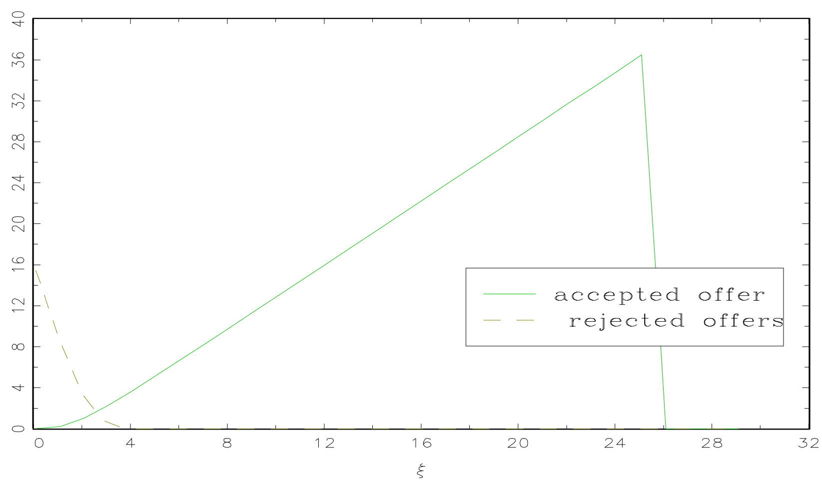

Figure 7: Jeffreys' Prior for ρ_{12} in the Roy Selection model