Focused Econometric Estimation: The interplay between the Bayesian and frequentist approaches

Andrés Ramírez Hassan* Manuel Correa Giraldo†

June 6, 2017

Abstract

Central to many econometric inferential situations is the estimation of non-linear functions of parameters. The mainstream in econometrics estimates these quantities based on a plug-in approach, where parameter estimates are just plugged in to the objective expressions without consideration of the main objective of the inferential situation. However, this approach suffers from many shortcomings, such as infinite moments and unbounded risks. Therefore, we introduce the Bayesian Minimum Expected Loss approach using generalized loss functions to estimate functions of parameters, and calculate their frequentist variability to avoid a prior sensitivity analysis. Simulation exercises show that our proposal outperforms competing alternatives in situations characterized by small sample sizes and noisy models. In addition, we observe in the applications that our approach gives lower standard errors than frequently used alternatives in these scenarios.

JEL Classification: C18, C13, C11.

Keywords: Bayesian Minimum Expected Loss, Frequentist variability, Functions of parameters, Parametric Bootstrap.

*Department of Economics, School of Economics and Finance, Universidad EAFIT, Medellín, Colombia; email: aramir21@eafit.edu.co

†Department of Economics, School of Economics and Finance, Universidad EAFIT, Medellín, Colombia; email: mcorre33@eafit.edu.co

1 Introduction

Central to many econometric inferential situations is the estimation of non-linear functions of the parameters. These functions might be elasticities, forecasts, impulse responses, marginal effects, optimal quantities, or structural parameters, among others. The mainstream in econometrics estimates these quantities based on a *plug-in* approach, where parameter estimates are just plugged in to the objective expressions without consideration of the main objective of the inferential situation. The popularity of this approach is based on the asymptotic properties of the delta method. However, this approach suffers from many shortcomings, such as infinite moments and unbounded risks.

Models have a purpose. So, we should optimally design a model's estimation for its purpose. This article is concerned with the estimation of non-linear functions of parameters, but deviates from the mainstream in that we focus the estimation process directly on the quantities of interest. This idea has been used by Claeskens and Hjort (2003) and Hansen (2005) for model selection, and DiTraglia (2016) for selecting moment conditions in the generalized method of moments (GMM).

We extend the idea of the Bayesian Minimum Expected Loss (MELO) approach introduced by Zellner (1978), whose main theoretical developments and applications were confined to structural econometric models. We introduce it for problems where the main concerns of the inference are functions of the parameters. In particular, we follow a decision theoretic framework where the posterior expected value of a generalized quadratic loss function that depends explicitly on the function of interest is minimized.

However, we will see that from a Bayesian perspective the MELO point estimates are exact, so we follow the idea of Efron (2012, 2015), who proposes to estimate the frequentist variability of the Bayesian estimates using posterior chains. Therefore, we avoid a sensitivity analysis of the choice of prior or hierarchical prior structure, and as a consequence their extra computational burden. However, this procedure is based on sufficient statistics, which might not be available. In that case, we propose a parametric bootstrap, at the cost of its extra computational burden, but improving the inferential framework by taking into account the biases, asymmetries, and support restrictions of the MELO point estimate. So, we propose an interplay between the Bayesian and frequentist approaches, which has been deeply studied and proposed by Good (1992); Bayarri and Berger (2004); Kass (2011); Efron (2012) and Ramírez Hassan (2017).

We perform different simulation exercises to compare the finite sample properties of our proposal with competing alternatives. We find that our proposal obtains better outcomes regarding point estimation of functions of parameters than competing alternatives; especially in settings characterized by noisy models and small sample sizes. In addition, we apply our proposal to real datasets, finding that MELO is more efficient than other alternatives.

The MELO approach has its foundation in statistical decision theory (Wald, 1945, 1947), which initially was advocated in econometrics by Marschak (1960) and Drèze (1974). It was introduced in econometrics by Zellner (1978), who analyzed reciprocals and ratios of parameters, and structural parameters in econometric models. He showed for these cases that the MELO estimator has, at least, finite first and second moments, and as a consequence finite risk with respect to a generalized quadratic loss function. On the other hand, common estimators like indirect least squares (ILS), two stage least squares (2SLS), limited information maximum likelihood (LIML), three stage least squares (3SLS) and full information maximum likelihood (FIML) have infinite moments and infinite risks using quadratic loss functions. Further, Zellner and Park (1979) approximate the small sample moments and risk functions of the MELO estimators, and compared them with other estimators. Zellner and Park (1980) found that coefficient estimates of structural parameters using MELO are matrix weighted averages of direct least squares (DLS) and 2SLS. Park (1982) showed through simulation exercises that for structural parameters, the MELO estimates have more bias than 2SLS. However, MELO outperforms 2SLS in criteria like mean squared error (MSE) and mean absolute error (MAE). Swamy and Mehta (1983) analyzed the requirements of prior distributions for reduced form parameters associated with the MELO estimator in undersized sample conditions, that is, situations where the number of exogenous variables in simultaneous equations models exceeds the sample size. They found that the conditions for existence of the FIML estimator are more demanding than the conditions to obtain the MELO. Diebold and Lamb (1997) used the MELO approach to do an interesting application related to the response of agricultural supply to movements in expected price. They argued that the large variability of previous estimates associated with this phenomenon is due to the infinite moments and multimodal distributions of common frequentist estimators. In contrast, the MELO estimator has at least finite first and second moments; however, it also may exhibit multimodal distributions. Finally, Zellner (1998) introduces the Bayesian Method of Moments, and related it to the MELO, extending the approach to cases where we have only moment conditions for our inferential problem. He present the resuts of simulation exercises that show that Bayesian estimators perform better than popular frequentist estimators.

This paper is structured as follows. The next section develops the theoretical framework. Section 3 exhibits the outcomes of the simulation exercises. Section 4 presents the main findings in our applications. Finally, we make some concluding remarks.

2 Theoretic framework

Let $(Y_1, Y_2, \ldots, Y_M)$ be random vectors with joint probability density function $f_Y(\mathbf{y}|\theta)$, suppose $Y \in \mathcal{R}^{N\times M}, \theta \in \mathcal{R}^L$, and let $\pi(\theta)$ be the prior distribution. Then

$$\pi(\theta|\mathbf{y}) = \frac{f_Y(\mathbf{y}|\theta)\pi(\theta)}{f_Y(\mathbf{y})}$$

where $f_Y(\mathbf{y}) = \int_\Theta f_Y(\mathbf{y}|\theta)\pi(\theta)d\theta$.

Suppose that the main concern of the econometric inference is $\omega = \mathbf{g}(\theta) : \mathcal{R}^L \to \mathcal{R}^K$, $K \leq L$, that is, $\omega = (\omega_1, \omega_2, \ldots \omega_K)^T = (g_1(\theta), g_2(\theta), \ldots, g_K(\theta))^T$, $g_k(\theta) : \mathcal{R}^L \to \mathcal{R}, k = 1, 2, \ldots, K$, such that

$$\begin{aligned} \gamma = h(\theta) : \mathcal{R}^L &\longrightarrow \mathcal{R}^L \\ \theta &\longmapsto (\mathbf{g}(\theta), f(\theta)) \end{aligned}$$

is a one-to-one continuously differentiable transformation for some nuisance transformation $\psi = f(\theta) : \mathcal{R}^L \rightarrow \mathcal{R}^{L-K}$.
Our view is that such an inferential problem should be directly tackled focusing on the functions of interest. So, we propose for this inferential problem the posterior Bayesian action that minimizes the posterior expected value of a generalized quadratic loss function focused on $\mathbf{g}(\theta)$, that is,

$$\min_{\hat{\omega}\in\mathcal{R}^K} E_{\pi_\theta(\theta|\mathbf{y})}\left\{\mathcal{L}(\mathbf{g}(\theta),\hat{\omega})\right\} = \min_{\hat{\omega}\in\mathcal{R}^K} \int_\Theta \left\{\mathcal{L}(\mathbf{g}(\theta),\hat{\omega})\right\} \pi_\theta(\theta|\mathbf{y})d\theta$$

where $\mathcal{L}(\mathbf{g}(\theta),\hat{\omega}) = (\mathbf{g}(\theta)-\hat{\omega})^T\mathbf{Q}(\theta)(\mathbf{g}(\theta)-\hat{\omega})$, $\mathbf{Q}(\theta)$ is a diagonal positive definite matrix.

Theorem 2.1. *The posterior Bayesian action, that is, the Minimum Expected Loss estimate, associated with* $\mathcal{L}(\mathbf{g}(\theta),\hat{\omega})$ *is*

$$\begin{aligned}\hat{\omega}^*(\mathbf{y}) &= \left[E_{\pi_\theta(\theta|\mathbf{y})}\mathbf{Q}(\theta)\right]^{-1} E_{\pi_\theta(\theta|\mathbf{y})}\left[\mathbf{Q}(\theta)\mathbf{g}(\theta)\right] \\ &= \left[\int_\Theta \mathbf{Q}(\theta)\pi_\theta(\theta|\mathbf{y})d\theta\right]^{-1}\left[\int_\Theta \mathbf{Q}(\theta)\mathbf{g}(\theta)\pi_\theta(\theta|\mathbf{y})d\theta\right]\end{aligned} \tag{1}$$

provided that $\left[E_{\pi_\theta(\theta|\mathbf{y})}\mathbf{Q}(\theta)\right]$ *is a positive definite finite nonsingular matrix,* $E_{\pi_\theta(\theta|\mathbf{y})}\left[\mathbf{Q}(\theta)\mathbf{g}(\theta)\right] < \infty$*, that is, the integral converges in each of its vector elements, and integration and differentiation can be interchanged.*

Proof in 6.1.
Observe that our MELO estimate is a weighted average of $\mathbf{g}(\theta)$, whose weights are given by $\left[\int_\Theta \mathbf{Q}(\theta)\pi_\theta(\theta|\mathbf{y})d\theta\right]^{-1}\mathbf{Q}(\theta)$. These weights implicitly depend on the probability associated with each θ in their parameter space as well as their magnitude. When $\mathbf{Q}$ does not depend on θ, which implies equal weight to each θ, the Minimum Expected Loss estimate is the posterior mean, that is, $\hat{\omega}^*(\mathbf{y}) = E_{\pi_\theta(\theta|\mathbf{y})}\mathbf{g}(\theta)$.

One great advantage of the MELO estimates is that they can be easily calculated from the draws of the posterior distributions, $\theta_s \sim \pi_\theta(\theta|\mathbf{y})$, and given $S\rightarrow\infty$, $\frac{1}{S}\sum_{s=1}^S \mathbf{Q}(\theta_s) \xrightarrow{p} E_{\pi_\theta(\theta|\mathbf{y})}\mathbf{Q}(\theta)$ and $\frac{1}{S}\sum_{s=1}^S \mathbf{Q}(\theta_s)\mathbf{g}(\theta_s) \xrightarrow{p} E_{\pi_\theta(\theta|\mathbf{y})}\left[\mathbf{Q}(\theta)\mathbf{g}(\theta)\right]$ by the law of the large numbers, then

$$\hat{\omega}^*_S(\mathbf{y}) = \left[\frac{1}{S}\sum_{s=1}^S \mathbf{Q}(\theta_s)\right]^{-1}\left[\frac{1}{S}\sum_{s=1}^S \mathbf{Q}(\theta_s)\mathbf{g}(\theta_s)\right] \tag{2}$$

converges in probability to $\hat{\omega}^*$ by Slutsky's theorem.

From a frequentist perspective, there are many situations when the posterior Bayesian actions are equal to the Bayes rules, that is, the estimators that minimize the Bayes risk, $r(\pi_\theta,\hat{\omega}) = \int_\Theta\int_Y \mathcal{L}(\mathbf{g}(\theta),\hat{\omega})f_Y(\mathbf{y}|\theta)dy\,\pi(\theta)d\theta$. However, there are situations where the Bayes risk is infinite, for instance using improper priors in conjunction general quadratic loss functions, and as a consequence, the Bayes rule does not exist. Nevertheless, it is still possible to obtain the posterior Bayesian action.

We do not use a hierarchical structure in our Bayesian formulation, so our MELO point estimate is exact from a Bayesian perspective. However, we can estimate its frequentist variability (Efron, 2012, 2015); avoiding a sensitivity analysis of the choice of priors, or hierarchical Bayesian models, both imposing an extra computational burden, at the cost of requiring sufficient statistics. To accomplish this task we have the following result.

Theorem 2.2. *If $\hat{\theta}(\mathbf{y}) \in \mathcal{R}^P$ is a sufficient statistic for $f_Y(\mathbf{y}|\theta)$, then*

$$\hat{\omega}^*(\mathbf{y}) = \hat{\omega}^*(\hat{\theta}(\mathbf{y})) \tag{3}$$

where $\hat{\omega}^*(\hat{\theta}(\mathbf{y})) = \left[E_{\pi_\theta(\theta|\hat{\theta}(\mathbf{y}))}\mathbf{Q}(\theta)\right]^{-1} E_{\pi_\theta(\theta|\hat{\theta}(\mathbf{y}))}\left[\mathbf{Q}(\theta)\mathbf{g}(\theta)\right]$.

Proof: See 6.2.

Equality 3 shows that our MELO estimate can be obtained from the posterior distribution associated with the data or its sufficient statistic. The resulting data reduction helps to estimate the frequentist variability of the MELO.

Setting

$$\alpha_{\hat{\theta}(\mathbf{y})}(\theta) = \nabla_{\hat{\theta}(\mathbf{y})} log f(\hat{\theta}(\mathbf{y})|\theta) = \left(\frac{\partial}{\partial\hat{\theta}(\mathbf{y})_1} log f(\hat{\theta}(\mathbf{y})|\theta), \cdots, \frac{\partial}{\partial\hat{\theta}(\mathbf{y})_P} log f(\hat{\theta}(\mathbf{y})|\theta)\right) \tag{4}$$

we have the following useful result.

Lemma 2.3. *Given $\mathbf{Q}(\theta)$ and $\mathbf{g}(\theta)$, the gradient of $\hat{\omega}^*(\hat{\theta}(\mathbf{y}))$ is*

$$\begin{aligned}\nabla_{\hat{\theta}(\mathbf{y})}\hat{\omega}^*(\hat{\theta}(\mathbf{y})) = &\left\{E_{\pi_\theta(\hat{\theta}(\mathbf{y}))}[\mathbf{Q}(\theta)]\right\}^{-1} \\ &\times\left\{E_{\pi_\theta(\hat{\theta}(\mathbf{y}))}[(\mathbf{Q}(\theta)\mathbf{g}(\theta))\otimes\alpha_{\hat{\theta}(\mathbf{y})}(\theta)] - \left[E_{\pi_\theta(\hat{\theta}(\mathbf{y}))}[\mathbf{Q}(\theta)\otimes\alpha_{\hat{\theta}(\mathbf{y})}(\theta)]\right][\hat{\omega}\otimes I_P]\right\}\end{aligned} \tag{5}$$

where I_P is the identity matrix of order P, and the operator $\otimes$ denotes the Kronecker product.

See the proof in 6.3.

Corollary 2.4. *When $\mathbf{Q}(\theta)$ and $\mathbf{g}(\theta)$ are in $\mathcal{R}$, then*

$$\begin{aligned}\nabla_{\hat{\theta}(\mathbf{y})}\hat{\omega}^*(\hat{\theta}(\mathbf{y})) = &\frac{E_{\pi_\theta(\theta|\hat{\theta}(\mathbf{y}))}[Q(\theta)g(\theta)\alpha_{\hat{\theta}(\mathbf{y})}(\theta)|\hat{\theta}(\mathbf{y})]}{E_{\pi_\theta(\theta|\hat{\theta}(\mathbf{y}))}[Q(\theta)|\hat{\theta}(\mathbf{y})]} \\ &-\frac{E_{\pi_\theta(\theta|\hat{\theta}(\mathbf{y}))}[Q(\theta)g(\theta)|\hat{\theta}(\mathbf{y})]E_{\pi_\theta(\theta|\hat{\theta}(\mathbf{y}))}[Q(\theta)\alpha_{\hat{\theta}(\mathbf{y})}(\theta)|\hat{\theta}(\mathbf{y})]}{(E_{\pi_\theta(\theta|\hat{\theta}(\mathbf{y}))}[Q(\theta)|\hat{\theta}(\mathbf{y})])^2}\end{aligned} \tag{6}$$

See the proof in 6.4.

Lemma 2.3 allows calculating the frequentist variability of the MELO estimate 1 through the delta method.

Theorem 2.5. *Setting $\hat{\theta}(\mathbf{y}) \sim (\mu_\theta, \Sigma_\theta)$, the frequentist covariance matrix of $\hat{\omega}^*(\mathbf{y})$ is*

$$Var(\hat{\omega}^*(\mathbf{y})) = Var(\hat{\omega}^*(\hat{\theta}(\mathbf{y}))) \approx \nabla_\theta \hat{\omega}^*(\hat{\theta}) \Sigma_{\hat{\theta}} \nabla_\theta \hat{\omega}^*(\hat{\theta})^T \tag{7}$$

provided that $N \to \infty$, $\hat{\theta} \xrightarrow{p} \theta$.

See the proof in 6.5.

The setting of our formulation establishes Theorem 2.1 as an optimal point estimate for functions of parameters. In the case that an analytical solution does not exist, we can use draws of the posterior distributions to obtain the estimates (Equation 2).

Theorem 2.5 allows obtaining the frequentist variance of our Bayesian estimate, provided a sufficient statistic.

In the case that sufficient statistics are not available, we propose to calculate the frequentist variability of the optimal Bayesian point estimate through a parametric bootstrap. This comes at the cost of an extra computational burden, but it improves the inferential framework taking into account biases, asymmetries, and support restrictions.

Algorithm A1 Parametric Bootstrap

1: Draw $\mathbf{y}_s^*$, $s = 1, 2, \ldots, S$ from $f_Y(\mathbf{y}|\hat{\theta})$
2: For each $\mathbf{y}_s^*$ calculate $\hat{\omega}_s^*(\mathbf{y}_s^*)$
3: Use $\{\hat{\omega}_1^*, \hat{\omega}_2^*, \ldots, \hat{\omega}_S^*\}$ to perform a statistical inference, such as means and standard deviations.

3 Simulation exercises

3.1 Optimal input

An economic problem

We consider a very simple problem where a firm is interested in finding the level of input (x) that maximizes its profit, where the production function is quadratic, that is, $y = \beta_1 x + \beta_2 x^2$. So, the problem is

$$\max_x \Pi(x) = \max_x IT(x) - CT(x) = \max_x p(\beta_1 x + \beta_2 x^2) - CF - wx$$

where p is the product's price, CF represents the fixed costs, and w is the input's price.

Then the optimal input is given by

$$x^{Opt} = \frac{1}{2\beta_2}\left[\frac{w}{p} - \beta_1\right] \tag{8}$$

This implies that the optimal production and profit are $y^{Opt} = \frac{1}{2\beta_2}\left[\left(\frac{w}{p}\right)^2 - \beta_1^2\right]$ and $\Pi^{Opt} = \frac{\beta_1}{2\beta_2}\left[w - \beta_1 p\right] - CF$, respectively.

An inferential problem

Suppose that the decision problem is to find the optimal level of input (Equation 8), that is, $\mathbf{g}(\theta) = \omega(\beta_1, \beta_2) = x^{Opt}$.

We can exploit the variability between y_i and x_i in the product function, the variability between x_i and w_i/p_i or y_i and w_i/p_i in the optimal input or production functions, or the variability between Π_i, w_i and p_i in the optimal profit function to obtain estimates of β_1 and β_2. The choice depends on assumptions regarding the rationality of the firms as well as the availability of the data.

We propose to formulate the mean deviation model associated with the production function to obtain the parameter estimates $\beta = [\beta_1 \; \beta_2]'$. In particular, $y_i - \bar{y} = \beta_1(x_i - \bar{x}) + \beta_2(x_i^2 - \overline{x^2}) + u_i$, where $\bar{y} = (1/N)\sum_{i=1}^N y_i$, $\bar{x} = (1/N)\sum_{i=1}^N x_i$, $\overline{x^2} = (1/N)\sum_{i=1}^N x_i^2$ and $u_i \sim \mathcal{N}(0, \sigma^2)$, $i = 1, 2, \ldots, N$.

The likelihood function of this model is

$$f(\beta, \sigma | y, X) \propto \sigma^{-N} exp\left\{-\left[vs^2 + \left(\beta - \hat{\beta}\right)^T X^T X \left(\beta - \hat{\beta}\right)\right]/2\sigma^2\right\}$$

where X is the design matrix, $q = dim\{\beta\}$, $v = N - q$, $\hat{\beta} = (X^T X)^{-1} X^T y$ and $vs^2 = (y - X\hat{\beta})^T(y - X\hat{\beta})$. $\hat{\beta}$ and s^2 are sufficient independent statistics, such that $\hat{\beta} \sim \mathcal{N}_q(\beta, \sigma^2(X^T X)^{-1})$ and $s^2 \sim \left(\frac{\sigma^2}{N-q}\right)\chi^2_{N-q}$. This implies

$$\Sigma_{\hat{\beta}, s^2} = \begin{bmatrix} \sigma^2(X^T X)^{-1} & 0 \\ 0 & \frac{2\sigma^4}{N-q} \end{bmatrix} \tag{9}$$

and

$$\alpha_{\hat{\beta}, s^2} = \left[(1/\sigma^2)(\beta - \hat{\beta})^T(X^T X) \quad (1/s^2)((N-q)/2 - 1) - 1/2\right] \tag{10}$$

The *plug-in* estimator for the optimal input would be

$$\hat{\omega}^{plug} = \frac{1}{2\hat{\beta}_2}\left(\frac{w}{p} - \hat{\beta}_1\right) \tag{11}$$

In addition, the application of the delta method to estimate the variance would give as a result

$$\widehat{Var(\hat{\omega}^{plug})} = \frac{1}{4\hat{\beta}_2^2}\left[\widehat{Var(\hat{\beta}_1)} + 4(\hat{\omega}^{plug})^2\widehat{Var(\hat{\beta}_2)} + 4\hat{\omega}^{plug}\widehat{Cov(\hat{\beta}_1, \hat{\beta}_2)}\right] \tag{12}$$

On the other hand we can obtain the MELO estimate focusing directly on the inferential problem. We set $\epsilon = -\left(\frac{w}{p} - \beta_1\right) - 2\beta_2\hat{\omega}$ as the estimation error. Observe that if $\hat{\omega}$ is equal to x^{Opt}, the estimation error is equal to 0.

The generalized loss function for this problem is given by

$$\begin{aligned}\mathcal{L}(\mathbf{g}(\theta), \hat{\omega}) = \epsilon^2 &= \left[\left(\frac{w}{p} - \beta_1\right) - 2\beta_2\hat{\omega}\right]^2 \\ &= 4\beta_2^2(\omega - \hat{\omega})^2\end{aligned}$$

where $\omega = \mathbf{g}(\theta) = \frac{1}{2\beta_2}\left(\frac{w}{p} - \beta_1\right)$ and $\mathbf{Q}(\theta) = 4\beta_2^2$.

Theorem 2.1 implies that the MELO estimate is

$$\begin{aligned}
\hat{\omega}^* &= \frac{\frac{w}{p}E(\beta_2) - E(\beta_1\beta_2)}{2E(\beta_2^2)} \\
&= \hat{\omega}^{plug}\left[\frac{1 - \left(\frac{1}{w/p - E(\beta_1)}\right)\left(Cov(\beta_1, \beta_2)/E(\beta_2)\right)}{\left(1 + Var(\beta_2)/E(\beta_2)^2\right)}\right]
\end{aligned} \tag{13}$$

Using the following diffuse prior $p(\beta, \sigma) \propto 1/\sigma$, $0 < \sigma < \infty$ and $-\infty < \beta_l < \infty$, $l = \{1, 2\}$, then the marginal posterior pdf for β has the form of a multivariate Student-t (Zellner, 1996):

$$p(\beta|y, X) \propto \left\{vs^2 + \left(\beta - \hat{\beta}\right)^T X^T X \left(\beta - \hat{\beta}\right)\right\}^{-(v+q)/2}$$

which implies that the mean of β is $\hat{\beta}$ and its covariance matrix is $(X^TX)^{-1}vs^2/(v-2)$, $v > 2$.

We can use the previous expressions to calculate our MELO proposal (Equation 13), and Equations 9 and 10 to obtain the frequentist variance of the MELO estimate.

Properties of MELO

We illustrate the properties of the MELO estimate for this simple example. In particular, we know that in this setting, $Cov(\beta_1, \beta_2) < 0$ (Uriel et al., 1997), this implies that if $\hat{\omega}^{plug} > 0$, which is a sensible outcome for an optimal input, then $\hat{\omega}^* > 0$. In this situation, $\hat{\omega}^{plug} > \hat{\omega}^*$ when $Var(\beta_2)E(\beta_2)^2 > -\left(\frac{1}{w/p - E(\beta_1)}\right)\left(\frac{Cov(\beta_1,\beta_2)}{E(\beta_2)}\right) > 0$. On the contrary, $\hat{\omega}^* > \hat{\omega}^{plug}$. In the case that $\hat{\omega}^{plug} < 0$, which does not make sense, then there is a chance that $\hat{\omega}^* > 0$; this happens when $\left(\frac{1}{w/p - E(\beta_1)}\right)\left(\frac{Cov(\beta_1,\beta_2)}{E(\beta_2)}\right) > 1$. So, the probability that the optimal input is positive is higher using the MELO than the *plug-in* approach.

Expanding Equation 13,

$$\hat{\omega}^* = \frac{1}{2E(\beta_2)}\left(w/p - E(\beta_1)\right)\left(\frac{1 - \left(1/(w/p - E(\beta_1))\right)\left(Cov(\beta_1, \beta_1)/E(\beta_2)\right)}{1 + Var(\beta_2)/E(\beta_2)^2}\right)$$

we can write

$$\begin{aligned}
\sqrt{n}\left(\hat{\omega}^* - \omega\right) &= \sqrt{n}\left[\frac{1}{2E(\beta_2)}\left(w/p - E(\beta_1)\right) - \omega\right] - \sqrt{n}\left[\frac{1}{2E(\beta_2)}\left(w/p - E(\beta_1)\right)\right] \\
&\times \left[\frac{2(1/(w/p - E(\beta_1)))^2 Cov(\beta_1, \beta_2) + 2(1/(w/p - \beta_1))Var(\beta_2)/E(\beta_2)^2}{1 + Var(\beta_2)/E(\beta_2)^2}\right]
\end{aligned} \tag{14}$$

Rewriting the first term on the right side of Equation 14,

$$
\begin{aligned}
\sqrt{n}\left[\frac{1}{2E(\beta_2)}\left(w/p - E(\beta_1)\right) - \omega\right] &= \sqrt{n}\left[\frac{1}{2E(\beta_2)}\left(w/p - E(\beta_1)\right) - \frac{1}{2\beta_2}\left(w/p - \beta_1\right)\right] \\
&= \sqrt{n}\left[\frac{w/p}{2}\left(\frac{1}{E(\beta_2)} - \frac{1}{\beta_2}\right) - \frac{1}{2}\left(\frac{E(\beta_1)}{E(\beta_2)} - \frac{\beta_1}{\beta_2}\right)\right]
\end{aligned}
$$

Zellner (1978) shows that $\sqrt{n}\left(\frac{1}{E(\beta_2)} - \frac{1}{\beta_2}\right)$ and $\sqrt{n}\left(\frac{E(\beta_1)}{E(\beta_2)} - \frac{\beta_1}{\beta_2}\right)$ converge in distribution to a normal distribution. So, applying the continuous mapping theorem, $\sqrt{n}\left[\frac{1}{2E(\beta_2)}\left(w/p - E(\beta_1)\right) - \omega\right]$ converges in distribution to a normal.

Regarding the second term on the right hand side of Equation 14, we have by Slustsky's theorem that

$$
\begin{aligned}
plim\ \hat{\omega}^* &= plim\left[\frac{1}{2E(\beta_2)}\left(w/p - E(\beta_1)\right)\right] \\
&= \frac{w/p}{2} plim\left(\frac{1}{E(\beta_2)}\right) - \frac{1}{2} plim\left(\frac{E(\beta_1)}{E(\beta_2)}\right) \\
&= \frac{w/p}{2}\frac{1}{plimE(\beta_2)} - \frac{1}{2}\frac{plimE(\beta_1)}{plimE(\beta_2)} \\
&= \frac{w/p}{2}\left(\frac{1}{\beta_2}\right) - \frac{1}{2}\left(\frac{\beta_1}{\beta_2}\right) \\
&= \omega
\end{aligned}
$$

Given that $plimE(\beta_2) \neq 0$ and $plim\sqrt{n}\,Cov(\beta_1, \beta_2) = plim\sqrt{n}\,Var(\beta_2) = 0$, then $\left[\frac{2(1/(w/p - E(\beta_1)))^2 Cov(\beta_1,\beta_2) + 2(1/(w/p - \beta_1))Var(\beta_2)/E(\beta_2)^2}{1 + Var(\beta_2)/E(\beta_2)^2}\right]$ converges in probability to 0. As a consequence, the second term on the right hand of expression 14 converges in probability to 0. Then by the transformation theorem, $\sqrt{n}(\hat{\omega}^* - \omega)$ converges in distribution to a normal distribution with mean equal to 0 and variance equal to $Var(\hat{\omega}^{plug})$. As a consequence, the MELO has the same asymptotic properties as the *plug-in* approach when it is seen as an estimator.

Simulation exercises

We set the mean deviation problem, $y_i - \bar{y} = 1.5(x_i - \bar{x}) - 0.002(x_i^2 - \overline{x^2}) + u_i$, where $x_i \sim \mathcal{N}(187.5, 70^2)$ and $u_i \sim \mathcal{N}(0, \sigma_u^2)$ such that σ_u^2 generates different degrees of signal to noise models $\{0.1, 1, 5, 20\}$. In addition, we set the input and output prices equal to \$3,000 and \$4,000, respectively. This implies $x^{Opt} = 187.5$.

We perform 1,000 simulation exercises using different sample sizes (20, 50 and 500), and calculate the Mean Squared Error (MSE) and the Mean Absolute Error (MAE) for the *plug-in* approach, and the MELO using the analytical solution (Equation 1), which is available in this setting, and the computational strategy of drawing from the posterior distribution

(Equation 2 using 10,000 iterations from a Student's t distribution).

We see from Table 1 that the MELO outperforms the *plug-in* approach in point estimates of the optimal input; especially in the presence of noisy models and small sample sizes. In addition, we observe that there is no meaningful difference between the analytical and computational solutions.

In particular, there is no clear pattern in the MSE and MAE in very noisy models as the sample size increases. However, we do observe that the MELO estimates outperform the *plug-in* approach in this situation. As the signal of the model improves, the MSE and MAE decrease as the sample size increases. The MSE and MAE from the MELO estimates (analytical and computational) are never worse than the *plug-in* estimates. However, we basically get the same outcomes using large sample sizes. This outcome follows from the previous asymptotic properties.

Table 1: Optimal input: Mean Errors

Signal/Noise	Method	Sample size	MSE	MAE
0.1	Plug-in	20	1,938,902.07	188.63
	Analytical MELO	20	1,155.06	13.25
	Computational MELO	20	1,153.36	13.24
	Plug-in	50	1,376,539.53	197.18
	Analytical MELO	50	5,610.28	15.70
	Computational MELO	50	5,571.76	15.67
	Plug-in	500	30,593,621.65	337.20
	Analytical MELO	500	3,808.44	16.84
	Computational MELO	500	3,807.83	16.85
1	Plug-in	20	426,199.12	146.97
	Analytical MELO	20	323.90	12.74
	Computational MELO	20	323.80	12.75
	Plug-in	50	13,846.97	40.31
	Analytical MELO	50	346.66	13.83
	Computational MELO	50	346.26	13.82
	Plug-in	500	124.33	7.46
	Analytical MELO	500	116.58	7.17
	Computational MELO	500	116.64	7.17
5	Plug-in	20	189.93	9.33
	Analytical MELO	20	112.92	8.21
	Computational MELO	20	112.88	8.21
	Plug-in	50	26.75	4.03
	Analytical MELO	50	26.05	3.97
	Computational MELO	50	26.05	3.97
	Plug-in	500	4.55	1.44
	Analytical MELO	500	4.54	1.43
	Computational MELO	500	4.54	1.43
20	Plug-in	20	7.06	2.10
	Analytical MELO	20	7.00	2.09
	Computational MELO	20	7.00	2.09
	Plug-in	50	1.61	0.99
	Analytical MELO	50	1.61	0.99
	Computational MELO	50	1.61	0.99
	Plug-in	500	0.28	0.36
	Analytical MELO	500	0.28	0.36
	Computational MELO	500	0.28	0.36

3.2 Odds ratio problem

Odds ratio

Setting y_i as a dichotomous variable $\{0, 1\}$ that is distributed as a Bernoulli process with parameter p, and assuming that the main interest is the Odds ratio, it follows that

$$\mathbf{g}(\theta) = \omega(p) = \frac{p}{1-p}$$

where $p = P(y = 1)$.

Inferential problem

The binary probit model can be used to tackle this situation, such that $p = P(y_i = 1) = \Phi(x_i^T \beta)$, where $\Phi(z)$ is the cumulative distribution function of the standard normal distribution evaluated at z.

This model can be written with latent variables as follows:

$$y_i^* = x_i^T \beta + u_i,\ u_i \sim \mathcal{N}(0,1) \tag{15}$$

$$y_i = \begin{cases} 0, & if \quad y_i^* \leq 0 \\ 1, & \quad\;\; y_i^* > 0 \end{cases} \tag{16}$$

The likelihood function is

$$f(\beta|y,x) = \prod_{i=1}^{N} \left(\Phi(x_i^T\beta)^{y_i} (1 - \Phi(x_i^T\beta))^{(1-y_i)} \right)$$

Observe that in this setting there are no sufficient statistics (Nelder and Wedderburn, 1972).

The *plug-in* estimator for the Odds ratio is

$$\hat{\omega}^{plug} = \frac{\Phi(x_i^T\hat{\beta})}{1 - \Phi(x_i^T\hat{\beta})}$$

And its variance, calculated by the delta method, is

$$\widehat{Var(\hat{\omega}^{plug})} = \frac{\Phi(x_i^T\hat{\beta})}{N\left[1 - \Phi(x_i^T\hat{\beta})\right]^3} \tag{17}$$

The loss function is given by

$$\mathcal{L}(\mathbf{g}(\theta), \hat{\omega}) = \epsilon^2 = \left[(1 - \Phi(x_i^T\beta))\hat{\omega} - \Phi(x_i^T\beta)\right]^2$$

where $\omega = \mathbf{g}(\theta) = \frac{\Phi(x_i^T\beta)}{1-\Phi(x_i^T\beta)}$ and $\mathbf{Q}(\theta) = (1 - \Phi(x_i^T\beta))^2$.

Theorem 2.1 implies that the MELO estimate is

$$\hat{\omega}^* = \frac{E\left[(1 - \Phi(x_i^T\beta))\Phi(x_i^T\beta)\right]}{E(1 - \Phi(x_i^T\beta))^2}$$

Note that if $\hat{p} \to 1$, then $1 - \hat{p} \to 0$, and so $\hat{\omega}^{plug} \to \infty$, while $\hat{\omega}^*$ can take indeterminate values of the form $0/0$.

According to Greenberg (2012), using the latent variables y_i^*, we can write the likelihood function as

$$\begin{aligned} f(y_i|y_i^*, \beta) &= [1(y_i = 0)1(y_i^* \leq 0) + 1(y_i = 1)1(y_i^* > 0)]\, \mathcal{N}_N(y^*|X\beta, I) \\ &= [1(y_i = 0)1(y_i^* \leq 0) + 1(y_i = 1)1(y_i^* > 0)] \\ &\times \exp\left\{ -\frac{1}{2}\left[vs^2 + \left(\beta - \hat{\beta}\right)^T X^T X \left(\beta - \hat{\beta}\right) \right] \right\} \end{aligned}$$

where $q = dim\{\beta\}$, $v = N - q$, $\hat{\beta} = (X^T X)^{-1} X^T y^*$ and $vs^2 = (y^* - X\hat{\beta})^T (y^* - X\hat{\beta})$. This implies that augmenting the observed binary data y with the latent variable y^*, $\hat{\beta}$ and vs^2 are sufficient statistics, so that

$$\Sigma_{\hat{\beta},s^2} = \begin{bmatrix} \sigma^2 (X^T X)^{-1} & 0 \\ 0 & \frac{2}{N-q} \end{bmatrix} \tag{18}$$

and

$$\alpha_{\hat{\beta},s^2} = \begin{bmatrix} (\beta - \hat{\beta})^T (X^T X) & (1/s^2)((N-q)/2 - 1) - 1/2 \end{bmatrix} \tag{19}$$

Assuming a normally distributed prior for β, the posterior distributions of β and y^* are

$$\pi(\beta, y^*|y) \propto \prod_{i=1}^{N} \{1(y_i = 0)1(y_i^* \leq 0) + 1(y_i = 1)1(y_i^* > 0)\} \mathcal{N}_N(y^*|X\beta, I)\mathcal{N}_q(\beta|\beta_0, B_0) \tag{20}$$

Therefore,

Algorithm A2 Bayesian Probit Model

1: Choose a starting value $\beta^{(0)}$

2: At the gth iteration, draw

$$y_i^* \sim \begin{cases} \mathcal{TN}_{(-\infty,0)}(x_i^T \beta^{(g-1)}, 1), & y_i = 0 \\ \mathcal{TN}_{(0,\infty)}(x_i^T \beta^{(g-1)}, 1), & y_i = 1 \end{cases}$$

3: $\beta^{(g)} \sim \mathcal{N}_q(\hat{\beta}^{(g)}, B_1)$, where $B_1 = (X^T X + B_0^{-1})$ and $\hat{\beta}^{(g)} = B_1(X^T y^{*(g)} + B_0^{-1}\beta_0)$.

Simulation exercises

Consider the following setting.

$$y_i^* = 0.5 + 0.8x_{1,i} - 1.2x_{2,i} + \mu_i \tag{21}$$

We simulate the data set x_1, x_2 and the stochastic errors from standard normal distributions, and perform 1,000 simulation exercises using four different sample sizes: 20, 50, 500 and 1,000.

Tables 2 and 3 show the mean errors of our simulation exercises. In particular, we perform two different evaluations for the Odds ratio, $x = (1, 1, 1)$ and $x = (1, 0, 0)$ using Algorithm A2 while setting $B_0 = 10,000\ diag\{1, 1, 1\}$ and $\beta_0 = [0, 0, 0]$ with 25,000 iterations and a burn-in equal to 5,000. We see from these tables that the range of variability of the different measures of the MELO approach is lower than for the *plug-in* approach. We observe that when the sample size is small, the differences are remarkable, especially when $x = (1, 1, 1)$, that is, when the data is less informative (noisy) due to regressors not being located in their population means ($x = (1, 0, 0)$). We obtain similar results for both approaches as the sample sizes increases.

Table 2: Odds ratio problem: Mean Errors for $x = (1, 1, 1)$

Mean Square Error								
Method	Sample size	Min	1st Qu.	Median	Mean	3rd Qu.	Max	Range
Plug-in*	20	0.0000	0.2484	0.7826	375,860,467,070,937,000.0000	1.8843	173,364,700,848,257,000,000.0000	173,364,700,848,257,000,000.0000
MELO**	20	0.0000	0.1683	0.5962	0.9860	1.1016	29.7273	29.7273
Plug-in	50	0.0000	0.0492	0.2476	74.5848	0.6757	73,222.6480	73,222.6480
MELO	50	0.0000	0.0536	0.2125	0.5777	0.5394	78.8489	78.8489
Plug-in	500	0.0000	0.0043	0.0223	0.0599	0.0706	1.4684	1.4684
MELO	500	0.0000	0.0049	0.0216	0.0556	0.0663	1.2967	1.2967
Plug-in	1,000	0.0000	0.0029	0.0134	0.0304	0.0355	0.5353	0.5353
MELO	1,000	0.0000	0.0034	0.0136	0.0291	0.0355	0.4880	0.4880
Mean Absolute Error								
Method	Sample size	Min	1st Qu.	Median	Mean	3rd Qu.	Max	Range
Plug-in*	20	0.0039	0.4984	0.8846	36,623,751.8166	1.3727	13,166,802,985.0931	13,166,802,985.0892
MELO**	20	0.0007	0.4102	0.7721	0.7900	1.0496	5.4523	5.4516
Plug-in	50	0.0005	0.2218	0.4976	1.0006	0.8220	270.5968	270.5963
MELO	50	0.0011	0.2315	0.4610	0.5525	0.7344	8.8797	8.8786
Plug-in	500	0.0004	0.0658	0.1493	0.1870	0.2658	1.2118	1.2114
MELO	500	0.0000	0.0697	0.1471	0.1822	0.2575	1.1387	1.1387
Plug-in	1,000	0.0001	0.0540	0.1157	0.1370	0.1884	0.7316	0.7315
MELO	1,000	0.0000	0.0579	0.1165	0.1346	0.1883	0.6985	0.6985
Mean Absolute Percentage Error								
Method	Sample size	Min	1st Qu.	Median	Mean	3rd Qu.	Max	Range
Plug-in*	20	0.0033	0.4249	0.7541	31,219,633.2935	1.1702	11,223,939,122.9163	11,223,939,122.9129
MELO**	20	0.0006	0.3497	0.6582	0.6734	0.8947	4.6478	4.6471
Plug-in	50	0.0004	0.1891	0.4241	0.8530	0.7007	230.6682	230.6677
MELO	50	0.0009	0.1974	0.3929	0.4710	0.6261	7.5694	7.5685
Plug-in	500	0.0003	0.0561	0.1273	0.1594	0.2266	1.0330	1.0326
MELO	500	0.0000	0.0594	0.1254	0.1553	0.2195	0.9707	0.9707
Plug-in	1,000	0.0001	0.0461	0.0986	0.1168	0.1606	0.6237	0.6236
MELO	1,000	0.0000	0.0493	0.0993	0.1147	0.1606	0.5955	0.5954

* We discard the "∞" values ** We discard the "NA" values and when $Plug-in$ takes "∞" values

Table 3: Odds ratio problem: Mean Errors for $x = (1, 0, 0)$

Mean Square Error								
Method	Sample size	Min	1st Qu.	Median	Mean	3rd Qu.	Max	Range
Plug-in*	20	0.0000	0.2868	1.2486	1,218,392,675,880,750.0000	5.0878	286,970,935,766,421,000.0000	286,970,935,766,421,000.0000
MELO**	20	0.0000	0.1521	1.0307	2.8656	2.8027	112.7688	112.7688
Plug-in	50	0.0000	0.1201	0.4335	5,986,756.9526	1.3330	5,608,827,328.7254	5,608,827,328.7254
MELO	50	0.0000	0.1052	0.3735	3.0819	1.1375	471.5700	471.5700
Plug-in	500	0.0000	0.0068	0.0374	0.0977	0.1151	2.7585	2.7585
MELO	500	0.0000	0.0069	0.0377	0.0933	0.1120	2.5165	2.5165
Plug-in	1,000	0.0000	0.0048	0.0205	0.0439	0.0524	0.5349	0.5349
MELO	1,000	0.0000	0.0047	0.0202	0.0429	0.0538	0.5100	0.5100
Mean Absolute Error								
Method	Sample size	Min	1st Qu.	Median	Mean	3rd Qu.	Max	Range
Plug-in*	20	0.0018	0.5355	1.1174	2,622,113.0994	2.2556	535,696,682.6166	535,696,682.6148
MELO**	20	0.0052	0.3900	1.0152	1.2250	1.6741	10.6193	10.6141
Plug-in	50	0.0008	0.3465	0.6584	100.5325	1.1546	74,892.1046	74,892.1038
MELO	50	0.0004	0.3243	0.6112	0.9356	1.0665	21.7157	21.7153
Plug-in	500	0.0003	0.0823	0.1934	0.2381	0.3392	1.6609	1.6606
MELO	500	0.0003	0.0833	0.1942	0.2343	0.3346	1.5863	1.5860
Plug-in	1,000	0.0005	0.0693	0.1432	0.1671	0.2290	0.7313	0.7309
MELO	1,000	0.0003	0.0687	0.1420	0.1657	0.2321	0.7141	0.7138
Mean Absolute Percentage Error								
Method	Sample size	Min	1st Qu.	Median	Mean	3rd Qu.	Max	Range
Plug-in*	20	0.0008	0.2389	0.4986	1,170,013.2120	1.0065	239,033,242.4775	239,033,242.4767
MELO**	20	0.0023	0.1740	0.4530	0.5466	0.7470	4.7384	4.7361
Plug-in	50	0.0003	0.1546	0.2938	44.8586	0.5152	33,417.6096	33,417.6092
MELO	50	0.0002	0.1447	0.2727	0.4175	0.4759	9.6897	9.6896
Plug-in	500	0.0001	0.0367	0.0863	0.1063	0.1514	0.7411	0.7410
MELO	500	0.0001	0.0372	0.0867	0.1046	0.1493	0.7078	0.7077
Plug-in	1,000	0.0002	0.0309	0.0639	0.0746	0.1022	0.3263	0.3261
MELO	1,000	0.0002	0.0307	0.0634	0.0739	0.1035	0.3187	0.3185

* We discard the "∞" values **We discard the "NA" values and when $Plug - in$ takes "∞" values

3.3 Portfolio selection

A financial problem

One strategy for active portfolio management looks for finding the asset weights that maximize the Sharpe ratio, that is, the mean portfolio return per unit of risk.

$$\max_{\mathbf{w}\in\mathcal{R}^L} \frac{\mathbf{w}^T\tilde{\mu}}{(\mathbf{w}^T\tilde{\Sigma}\mathbf{w})^{1/2}} \qquad \text{s.t} \qquad \mathbf{w}^T\mathbf{1} = 1$$

where $\tilde{\mu}$ is the mean vector of the asset's excess returns in the investment period (say τ), $\tilde{\Sigma}$ is its covariance matrix, and $\mathbf{1}$ is a vector of ones.

The solution of the previous problem gives the well known tangent portfolio, that is,

$$\mathbf{w}^{Opt} = \frac{\tilde{\Sigma}^{-1}\tilde{\mu}}{\mathbf{1}^T\tilde{\Sigma}^{-1}\tilde{\mu}} \tag{22}$$

As we can see from Equation 22, the final aim of the inferential problem is a non-linear function of the parameters of the asset's excess returns.

Inferential problem

The standard financial literature assumes that the asset's excess returns are jointly normally distributed, i.e., $r_t \sim \mathcal{N}_d(\mu,\Sigma)$ for $t = 1,2,\ldots,T$, where the excess returns are serially independent.

Now put $\mathbf{R} = (\mathbf{r_1},\mathbf{r_2},\ldots,\mathbf{r_L})$ a $T\times L$ matrix of observations on L asset excess returns. Then we can write the following model for the excess returns:

$$\mathbf{R} = \mathbf{1}\mu^T + e$$

where $e = (e_1, e_2,\ldots,e_L)$ is an $T\times L$ matrix of unobserved random disturbances. The rows of e are independently distributed, which precludes any auto or serial correlation of disturbance terms, each with an L-dimensional normal distribution with zero mean vector and positive definite $L\times L$ covariance matrix Σ.

The likelihood of this model is

$$f(\mu,\Sigma|\mathbf{R}) \propto |\Sigma|^{-T/2} exp\left\{-\frac{1}{2}tr(S\Sigma^{-1}) - \frac{1}{2T}tr((\mu-\hat{\mu})(\mu-\hat{\mu})^T\Sigma^{-1})\right\}$$

where $\hat{\mu}$ is the sample mean vector and $S = (\mathbf{R}-\mathbf{1}\hat{\mu}^T)^T(\mathbf{R}-\mathbf{1}\hat{\mu}^T)$. $\hat{\mu}$ and S are sufficient statistics, such that $\hat{\mu} \sim \mathcal{N}_L(\mu,\Sigma)$ and $S \sim \mathcal{W}_L(T-1,\Sigma)$. $\hat{\mu}$ and $S/(T-1)$ are consistent estimators for μ and Σ. Then,

$$\alpha_{\hat{\mu},\hat{\Sigma}} = \left[(\mu-\hat{\mu})^T\Sigma^{-1} \quad vec\left(\left(\frac{T-1-L-1}{2}\right)S^{-1} - \frac{1}{2}\Sigma^{-1}\right)^T\right] \tag{23}$$

$$\Sigma_{\hat{\theta}} = \begin{bmatrix} \Sigma & 0 \\ 0 & \Sigma_S \end{bmatrix} \tag{24}$$

where $Var(S_{ij}) = (T-1)(\sigma_{ij}^2 + \sigma_{ii}\sigma_{jj})$ and $Cov(S_{ij}, S_{kl}) = (T-1)(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk})$.

The *plug-in* estimator for the tangent portfolio is

$$\hat{\mathbf{w}}^{plug} = \frac{\hat{\Sigma}^{-1}\hat{\mu}}{\mathbf{1}^T\hat{\Sigma}^{-1}\hat{\mu}} \tag{25}$$

On the other hand we can obtain the MELO estimate focusing directly on the inferential problem. We set $\epsilon = (\mathbf{1}^T\tilde{\Sigma}^{-1}\tilde{\mu})\hat{\omega} - \tilde{\Sigma}^{-1}\tilde{\mu}$ as the estimation error. Observe that if $\hat{\omega}$ is equal to $\mathbf{w}^{Opt}$, the estimation error is equal to 0.

Given $\mathbf{g}(\theta) = \frac{\tilde{\Sigma}^{-1}\tilde{\mu}}{\mathbf{1}^T\tilde{\Sigma}^{-1}\tilde{\mu}}$, the generalized loss function for this problem is given by $\mathcal{L}(\mathbf{g}(\theta), \hat{\omega}) = \epsilon^T\epsilon$ and $E(\mathcal{L}) = E_{\pi(\mu,\Sigma|\mathbf{R})}\epsilon^T\epsilon$, $\mathbf{Q}(\theta) = (\mathbf{1}^T\tilde{\Sigma}^{-1}\mu)^2$. Despite the fact that the expected value should be based on information up to the investment period $(T+\tau)$, we only have information up to T, so the expected value is conditioned on $\mathbf{R}$. However, informative priors can be based on experts' views of the investment period.

Theorem 2.1 implies that the MELO estimate is

$$\hat{\omega}^* = \frac{E_{\pi(\mu,\Sigma|\mathbf{R})}((\mathbf{1}^T\tilde{\Sigma}^{-1}\tilde{\mu})\tilde{\Sigma}^{-1}\tilde{\mu})}{E_{\pi(\mu,\Sigma|\mathbf{R})}(\mathbf{1}^T\tilde{\Sigma}^{-1}\tilde{\mu})^2} \tag{26}$$

Using the diffuse prior $\pi(\mu, \Sigma) = \pi(\mu)\pi(\Sigma) \propto |\Sigma|^{-(L+1)/2}$, the conditional posterior distribution for the mean vector of asset excess returns is $\mu|\Sigma, \mathbf{R} \sim \mathcal{N}_L(\hat{\mu}, \Sigma/T)$, and the marginal distribution for the covariance matrix is $\Sigma|\mathbf{R} \sim \mathcal{IW}_L(T-1, S)$ (Zellner, 1996). Therefore, we can use a Gibbs sampling algorithm to obtain a computational solution for our MELO estimate.[1]

Simulation exercises

We set $\mu_l \sim \mathcal{U}(-0.2, 0.2)$, $l = 1, 2, \dots, L$, and generate Σ such that it is semidefinite positive. We set four different scenarios of portfolio selection: $L = \{10, 25, 50, 100\}$ assets, and two sample sizes: $T = \{120, 240\}$ periods. We perform 100 simulations for each of the 8 settings, so that $\mathbf{R} \sim \mathcal{N}(\mu, \Sigma)$.

We estimate the sample mean and covariance matrix to calculate the optimal weights using the *plug-in* approach (Equation 25), and the Gibbs sampling algorithm with 1,000 iterations to calculate our MELO proposal (Equation 26). Then, we obtain the MSE and MAE using the population parameters (Equation 22), and the two estimators. We can see in

[1] Observe that following the *concentional* Bayesian portfolio selection, which is based on the predictive distribution of the excess returns in the investment period, we have $\tilde{\mu} = \hat{\mu}$ and $\tilde{\Sigma} = \frac{(\tau+\frac{1}{T})(T-1)}{T+\tau-2-L}\hat{\Sigma}$. The term $\frac{(\tau+\frac{1}{T})(T-1)}{T+\tau-2-L}$ cancels out in Equation 26.

Tables 4 and 5 the outcomes of our simulation exercises. In particular, the mean of the MSE and MAE associated with the MELO is always lower than the *plug-in* approach; there are remarkable improvements of our proposal when the number of assets in the portfolio selection problem is small. In the latter cases, the range of variability in the MSE and MAE using the *plug-in* approach is enormous compared with the MELO approach.

Table 4: Tangency portfolio: Mean Squared Error

Method	Sample size	Min	1st Qu.	Median	Mean	3rd Qu.	Max	Range
Mean Squared Error								
Assets = 10								
Plug-in	120	0.0307	0.0951	0.2266	12.5700	0.9281	627.6000	627.5693
MELO	120	0.0126	0.0627	0.0947	0.1089	0.1261	0.2787	0.2661
Plug-in	240	0.0200	0.0680	0.1970	41.8200	1.2060	3,306.0000	3,305.9800
MELO	240	0.0185	0.0606	0.0767	0.0949	0.1138	0.4032	0.3847
Assets = 25								
Plug-in	120	0.0048	0.0116	0.0221	0.7194	0.1083	36.2200	36.2152
MELO	120	0.0026	0.0067	0.0097	0.0112	0.0141	0.0454	0.0428
Plug-in	240	0.0027	0.0057	0.0102	0.0573	0.0205	1.3780	1.3753
MELO	240	0.0024	0.0049	0.0072	0.0083	0.0106	0.0345	0.0321
Assets = 50								
Plug-in	120	0.0150	0.0298	0.0627	20.1600	0.2210	1,618.0000	1,617.9850
MELO	120	0.0133	0.0266	0.0320	0.0421	0.0480	0.1242	0.1109
Plug-in	240	0.0062	0.0191	0.0282	5.7330	0.1950	131.1000	131.0938
MELO	240	0.0079	0.0184	0.0213	0.0340	0.0377	0.1024	0.0945
Assets = 100								
Plug-in	120	1.5100*e*-05	4.7920*e*-05	7.4650*e*-05	1.2590*e*-04	1.2210*e*-04	1.8410*e*-03	0.0018
MELO	120	1.1900*e*-05	4.7780*e*-05	7.4800*e*-05	1.2380*e*-04	1.2110*e*-04	1.7300*e*-03	0.0017
Plug-in	240	2.2100*e*-06	5.8250*e*-06	8.8950*e*-06	1.3330*e*-05	1.6180*e*-05	6.3400*e*-05	0.0001
MELO	240	2.1900*e*-06	5.9950*e*-06	9.0150*e*-06	1.3290*e*-05	1.5970*e*-05	6.2100*e*-05	0.0001

Table 5: Tangency portfolio: Mean Absolute Error

Method	Sample size	Min	1st Qu.	Median	Mean	3rd Qu.	Max	Range
Mean Absolute Error								
Assets = 10								
Plug-in	120	0.1263	0.2402	0.3796	1.1380	0.7703	17.6600	17.5337
MELO	120	0.0966	0.2021	0.2483	0.2504	0.2799	0.4638	0.3672
Plug-in	240	0.0974	0.2094	0.3747	1.6680	0.8882	54.8500	54.7526
MELO	240	0.1012	0.1966	0.2160	0.2320	0.2606	0.4689	0.3677
Assets = 25								
Plug-in	120	0.0515	0.0847	0.1165	0.3164	0.2595	4.2000	4.1486
MELO	120	0.0380	0.0659	0.0762	0.0794	0.0905	0.1694	0.1314
Plug-in	240	0.0427	0.0590	0.0814	0.1230	0.1117	0.8916	0.8489
MELO	240	0.0380	0.0547	0.0653	0.0683	0.0785	0.1575	0.1195
Assets = 50								
Plug-in	120	0.0845	0.1255	0.1853	0.8541	0.3524	31.1100	31.0255
MELO	120	0.0866	0.1121	0.1226	0.1329	0.1454	0.2284	0.1418
Plug-in	240	0.0633	0.0997	0.1233	0.6502	0.3045	8.6480	8.5847
MELO	240	0.0624	0.0901	0.0992	0.1171	0.1269	0.2267	0.1643
Assets = 100								
Plug-in	120	0.0029	0.0053	0.0066	0.0075	0.0084	0.0325	0.0296
MELO	120	0.0026	0.0053	0.0065	0.0075	0.0084	0.0315	0.0290
Plug-in	240	0.0012	0.0019	0.0023	0.0026	0.0031	0.0061	0.0049
MELO	240	0.0012	0.0019	0.0023	0.0026	0.0031	0.0060	0.0049

3.4 Structural supply–demand model

An economic problem

Assume the following structural supply–demand model:

$$q_i^d = \beta_0 + \beta_1 p_i + \beta_2 z_{1i} + \mu_{di} \tag{27}$$
$$q_i^s = \alpha_0 + \alpha_1 p_i + \alpha_2 z_{2i} + \mu_{si} \tag{28}$$

where q_i^d and q_i^s are demand and supply functions, p_i is the price, z_{1i} and z_{2i} exogenous regressors, and μ_{di} and μ_{si} stochastic errors, $i = 1, 2, \dots, N$.

The equilibrium condition equates demand and supply, that is, $q^s = q^d$. the structural parameters are the main concern of the econometric inferential problem.

The inferential problem

Equation 27 cannot be directly estimated due to endogeneity issues. So, it is necessary to obtain the reduced form system

$$q_i = \pi_0 + \pi_1 z_{1i} + \pi_2 z_{2i} + e_{qi}$$
$$p_i = \gamma_0 + \gamma_1 z_{1i} + \gamma_2 z_{2i} + e_{pi}$$

which can be written as $\mathbf{Y} = \mathbf{X}B + \mathbf{U}$, where $\mathbf{Y} = [q \quad p]$, an $N \times 2$ matrix of observations on quantities and prices, $\mathbf{X}$ is an $N \times 3$ matrix of a vector of ones, and the two independent variables (z_1 and z_2), with rank 3, $B = [\pi \quad \gamma]$ is a 3×2 matrix of regressions parameters from the reduced form, and $\mathbf{U} = [e_q \quad e_p]$ is a $N \times 2$ matrix of unobserved stochastic errors. We assume that the rows of U are independently distributed, each with a 2-dimensional normal distribution with zero mean vector and positive definite 2×2 covariance matrix Σ.

The likelihood function of this system is

$$f(B, \Sigma|\mathbf{Y}, \mathbf{X}) \propto |\Sigma|^{-T/2} exp\left\{-\frac{1}{2}tr(\mathbf{S}\Sigma^{-1}) - \frac{1}{2}tr((B - \hat{B})^T(\mathbf{X}^T\mathbf{X})(B - \hat{B})\Sigma^{-1})\right\}$$

where $\hat{B} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$ a matrix of least squares quantities, and $\mathbf{S} = (\mathbf{Y} - \mathbf{X}\hat{B})^T(\mathbf{Y} - \mathbf{X}\hat{B})$. $\hat{B}$ and S are sufficient statistics, such that $vec(\hat{B}) = \begin{bmatrix}\hat{\pi}^T & \hat{\gamma}^T\end{bmatrix}^T \sim \mathcal{N}_6(vec(B), \Sigma \otimes (\mathbf{X}^T\mathbf{X})^{-1})$ and $\mathbf{S} \sim \mathcal{W}_2(N - 3, \Sigma)$. $\hat{B}$ and $\mathbf{S}/(N - 3)$ are consistent estimators for B and Σ.

Then,

$$\alpha_{\hat{\beta},\hat{\Sigma}} = \left[(\beta - \hat{\beta})^T(\Sigma^{-1} \otimes (\mathbf{X}^T\mathbf{X})) \quad vec\left(\left(\frac{N - 3 - 2 - 1}{2}\right)\mathbf{S}^{-1} - \frac{1}{2}\Sigma^{-1}\right)^T\right]$$

and

$$\Sigma_{\hat{\beta},\hat{\Sigma}} = \begin{bmatrix}\Sigma \otimes (\mathbf{X}^T\mathbf{X})^{-1} & 0 \\ 0 & \Sigma_S\end{bmatrix}$$

where $\beta = vec(B)$, $\hat{\beta} = vec(\hat{B})$, $Var(S_{ij}) = (N-3)(\sigma_{ij}^2 + \sigma_{ii}\sigma_{jj})$ and $Cov(S_{ij}, S_{kl}) = (N-3)(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk})$.

The relation between the structural parameters, which are the main concern of the econometric inferential problem, and the reduced form parameters is given by the following system of equations:

$$\begin{bmatrix}\beta_1\\ \beta_2\\ \alpha_1\\ \alpha_2\end{bmatrix} = \begin{bmatrix}\pi_2/\gamma_2\\ \pi_1 - \gamma_1\pi_2/\gamma_2\\ \pi_1/\gamma_1\\ \pi_2 - \gamma_2\pi_1\gamma_1\end{bmatrix} \tag{29}$$

There are different alternatives for obtaining the structural parameters from the reduced form. In this setting, which is an exactly identified model, the point estimates using the ILS (*plug-in* approach), 2SLS, or 3SLS, give the same results.

We set the vector of errors to be

$$\begin{bmatrix}\epsilon_1\\ \epsilon_2\\ \epsilon_3\\ \epsilon_4\end{bmatrix} = \begin{bmatrix}\gamma_2(\hat{\omega}_1 - \beta_1)\\ \gamma_2(\hat{\omega}_2 - \beta_2)\\ \gamma_1(\hat{\omega}_3 - \alpha_1)\\ \gamma_1(\hat{\omega}_4 - \alpha_2)\end{bmatrix}$$

The loss function is $\mathcal{L}(\mathbf{g}(\theta), \hat{\omega}) = \epsilon^T\epsilon$, which implies $\mathbf{Q} = diag(\gamma_2^2, \gamma_2^2, \gamma_1^2, \gamma_1^2)$. As a consequence, the MELO is given by the following set of simultaneous equations.

$$\begin{bmatrix}\hat{\omega}_1^*\\ \hat{\omega}_2^*\\ \hat{\omega}_3^*\\ \hat{\omega}_4^*\end{bmatrix} = \begin{bmatrix}E(\pi_2\gamma_2)/E(\gamma_2^2)\\ (E(\pi_1\gamma_2^2) - E(\gamma_1\gamma_2\pi_2))/E(\gamma_2^2)\\ E(\pi_1\gamma_1)/E(\gamma_1^2)\\ (E(\pi_2\gamma_1^2) - E(\gamma_1\gamma_2\pi_1))/E(\gamma_1^2)\end{bmatrix}$$

Observe that the different components of the MELO estimates are independent. This is due to the structure of the weighting matrix: it is a diagonal matrix. So, we can focus our effort on the specific structural parameters of interest.

Using the diffuse prior $\pi(B,\Sigma) = \pi(B)\pi(\Sigma) \propto |\Sigma|^{-(2+1)/2}$, the conditional posterior distribution for the mean vector is $vec(B)|\Sigma, \mathbf{Y}, \mathbf{X} \sim \mathcal{N}_6(vec(\hat{B}), \Sigma \otimes (\mathbf{X}^T\mathbf{X})^{-1})$, and the marginal distribution for the covariance matrix is $\Sigma|\mathbf{Y}, \mathbf{X} \sim \mathcal{IW}_2(N-3, S)$ (Zellner, 1996). Therefore, we can use a Gibbs sampling algorithm to obtain a computational solution of our MELO estimate.

Simulation exercises

We consider the following structural supply–demand model:

$$q_i^d = 0.2 - 0.8p_i + 1.5z_{1i} + \mu_{di} \tag{30}$$

$$q_i^s = -0.5 + 1.2p_i - z_{2i} + \mu_{si} \tag{31}$$

which implies the following reduced form model:

$$q_i = 0.35 + 0.75z_{1i} + 0.50z_{2i} + e_{qi} \tag{32}$$

$$p_i = -0.08 + 0.9z_{1i} - 0.4z_{2i} + e_{pi} \tag{33}$$

We simulate z_{1i} and z_{2i} from standard normal distributions, and the stochastic errors from the reduced system as independent variables with mean zero and standard deviation such that the signal to the noise ratio in the reduced equations are simultaneously equal to 0.1, 0.5, 1 and 5. We know that from a theoretical point of view this is a mistake since there is a correlation between the stochastic errors in the reduced form system. However, we follow this setting to have independence between the equations, and as a consequence we know that the marginal posterior distributions of each equation are independent multivariate Student's t distributions. This implies that

$$\begin{bmatrix}\hat{\omega}_1^*\\ \hat{\omega}_2^*\\ \hat{\omega}_3^*\\ \hat{\omega}_4^*\end{bmatrix} = \begin{bmatrix} E(\pi_2)E(\gamma_2)/(Var(\gamma_2)+(E(\gamma_2))^2)\\ E(\pi_1)-(E(\pi_2)(Cov(\gamma_1,\gamma_2)+E(\gamma_1)E(\gamma_2))/(Var(\gamma_2)+(E(\gamma_2))^2)\\ E(\pi_1)E(\gamma_1)/(Var(\gamma_1)+(E(\gamma_1))^2)\\ E(\pi_2)-(E(\pi_1)(Cov(\gamma_1,\gamma_2)+E(\gamma_1)E(\gamma_2))/(Var(\gamma_1)+(E(\gamma_1))^2)\end{bmatrix}$$

and so we can compare the analytical and computational versions of the MELO estimates with the frequentist competing alternative. In particular, we use 50,000 iterations for the Gibbs sampling algorithm used to calculate the computational MELO.

We can see in Table 6 the mean errors associated with 1,000 simulations exercises using five different sample sizes: 20, 50, 100, 1,000 and 20,000. We observe from this table the same pattern as in the previous simulations exercises. The MELO estimates outperform 2SLS in terms of point estimates, especially in situations characterized by noisy models and small sample sizes. However, we always get the same performance with a sample size equal to 20,000 or a signal to noise ratio equal to 5. The performance of the three approaches improves as the sample size increases as well as when the signal to noise increases. In general, the MELO estimates are never worse than those of the 2SLS, and the analytical and computational solutions have the same performance.

4 Applications

4.1 Experimental broiler input–output

This is the broiler input–output example presented by Judge et al. (1988). In particular, the average weight of an experimental lot of broilers and their corresponding levels of average feed consumption was tabulated over the time period in which they changed from baby chickens to mature broilers ready for market.

Given the setting of the optimal input problem in subsection 3.1, the dataset in Table 5.3 from Judge et al. (1988), and taking into account that broilers are 30 cents per pound and feed is 6 cents per pound, the optimal level of feed input is 13.74 with a standard deviation equal to 1.89 using the *plug-in* approach (Equations 11 and 12), whereas the optimal input

Table 6: Demand and supply model: Mean Errors.

			Mean Squared Error				Mean Absolute Error				Mean Absolute Percentage Error			
Signal/Noise	Method	Sample size	β_1	β_2	α_1	α_2	β_1	β_2	α_1	α_2	β_1	β_2	α_1	α_2
	2SLS		884.83	5,030.75	2,433,378.77	14,396,948.28	5.35	10.76	54.55	129.84	669.36%	717.06%	4545.69%	12983.79%
	Analytical MELO	20	0.80	5.70	1.56	6.05	0.80	1.91	1.17	1.95	99.56%	127.20%	97.44%	194.72%
	Computational MELO		0.81	5.75	1.57	6.10	0.80	1.91	1.17	1.95	99.80%	127.64%	97.45%	195.45%
	2SLS		638.92	544.20	14,826.33	47,907.91	5.07	5.07	9.17	13.27	633.62%	337.88%	763.99%	1326.56%
	Analytical MELO	50	0.78	2.48	1.42	2.96	0.79	1.25	1.08	1.36	98.17%	83.41%	89.60%	136.29%
	Computational MELO		0.78	2.50	1.42	2.98	0.79	1.25	1.07	1.36	98.26%	83.58%	89.57%	136.49%
	2SLS		429.43	633.86	201.59	145.63	4.87	5.07	4.17	3.63	609.32%	338.20%	347.14%	362.62%
0.1	Analytical MELO	100	0.75	1.55	1.22	1.66	0.76	1.00	0.98	1.02	95.43%	66.35%	81.74%	102.41%
	Computational MELO		0.75	1.55	1.23	1.66	0.76	1.00	0.98	1.03	95.49%	66.40%	81.70%	102.51%
	2SLS		107.71	84.95	3.44	1.20	2.20	1.78	0.74	0.53	274.66%	118.61%	61.67%	52.89%
	Analytical MELO	1000	0.31	0.31	0.25	0.22	0.44	0.44	0.40	0.37	55.02%	29.51%	33.17%	37.17%
	Computational MELO		0.31	0.31	0.25	0.22	0.44	0.44	0.40	0.37	55.01%	29.50%	33.19%	37.18%
	2SLS		0.03	0.03	0.02	0.02	0.14	0.12	0.11	0.10	17.72%	8.33%	9.02%	10.40%
	Analytical MELO	20000	0.03	0.02	0.02	0.02	0.14	0.12	0.11	0.10	17.24%	8.16%	8.89%	10.34%
	Computational MELO		0.03	0.02	0.02	0.02	0.14	0.12	0.11	0.10	17.24%	8.16%	8.89%	10.34%
	2SLS		889.82	864.22	731.46	140.04	3.43	2.74	3.44	1.77	429.33%	182.80%	286.66%	177.36%
	Analytical MELO	20	0.51	0.46	0.45	0.45	0.59	0.54	0.54	0.53	73.95%	36.26%	44.64%	53.17%
	Computational MELO		0.52	0.46	0.45	0.46	0.59	0.55	0.53	0.54	74.20%	36.42%	44.57%	53.54%
	2SLS		336.38	98.10	8.28	2.69	1.98	1.27	0.66	0.50	247.52%	84.92%	54.86%	49.68%
	Analytical MELO	50	0.32	0.25	0.22	0.22	0.45	0.40	0.37	0.37	56.46%	26.55%	30.76%	36.71%
	Computational MELO		0.32	0.25	0.23	0.22	0.45	0.40	0.37	0.37	56.60%	26.61%	30.87%	36.79%
	2SLS		5.94	2.47	0.26	0.16	0.68	0.52	0.36	0.30	85.36%	34.90%	30.27%	30.22%
0.5	Analytical MELO	100	0.19	0.15	0.15	0.12	0.35	0.31	0.30	0.27	43.14%	20.54%	25.36%	27.49%
	Computational MELO		0.19	0.15	0.15	0.12	0.35	0.31	0.30	0.28	43.23%	20.56%	25.39%	27.51%
	2SLS		0.03	0.02	0.02	0.01	0.12	0.11	0.10	0.08	15.62%	7.47%	8.28%	8.23%
	Analytical MELO	1000	0.02	0.02	0.02	0.01	0.12	0.11	0.10	0.08	15.22%	7.32%	8.23%	8.19%
	Computational MELO		0.02	0.02	0.02	0.01	0.12	0.11	0.10	0.08	15.22%	7.32%	8.23%	8.19%
	2SLS		0.00	0.00	0.00	0.00	0.03	0.02	0.02	0.02	3.47%	1.64%	1.80%	2.07%
	Analytical MELO	20000	0.00	0.00	0.00	0.00	0.03	0.02	0.02	0.02	3.46%	1.63%	1.79%	2.07%
	Computational MELO		0.00	0.00	0.00	0.00	0.03	0.02	0.02	0.02	3.46%	1.63%	1.79%	2.08%
	2SLS		47.94	13.73	0.39	0.21	1.01	0.71	0.42	0.33	125.70%	47.61%	34.82%	33.21%
	Analytical MELO	20	0.22	0.17	0.18	0.15	0.37	0.33	0.33	0.30	45.99%	21.75%	27.55%	29.93%
	Computational MELO		0.22	0.18	0.19	0.15	0.37	0.33	0.33	0.30	46.44%	21.88%	27.80%	30.06%
	2SLS		0.60	0.29	0.09	0.06	0.37	0.29	0.23	0.20	46.14%	19.17%	18.87%	19.63%
	Analytical MELO	50	0.13	0.08	0.08	0.06	0.28	0.23	0.21	0.19	35.07%	15.39%	17.64%	19.09%
	Computational MELO		0.13	0.08	0.08	0.06	0.28	0.23	0.21	0.19	35.16%	15.42%	17.65%	19.10%
	2SLS		0.08	0.05	0.04	0.03	0.22	0.18	0.17	0.14	27.46%	12.19%	13.76%	14.13%
1	Analytical MELO	100	0.07	0.05	0.04	0.03	0.20	0.17	0.16	0.14	25.51%	11.47%	13.28%	13.87%
	Computational MELO		0.07	0.05	0.04	0.03	0.20	0.17	0.16	0.14	25.53%	11.47%	13.28%	13.88%
	2SLS		0.01	0.00	0.00	0.00	0.06	0.05	0.05	0.04	7.66%	3.67%	4.13%	4.11%
	Analytical MELO	1000	0.01	0.00	0.00	0.00	0.06	0.05	0.05	0.04	7.60%	3.64%	4.12%	4.11%
	Computational MELO		0.01	0.00	0.00	0.00	0.06	0.05	0.05	0.04	7.60%	3.64%	4.12%	4.11%
	2SLS		0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	1.73%	0.82%	0.90%	1.04%
	Analytical MELO	20000	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	1.73%	0.82%	0.90%	1.04%
	Computational MELO		0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	1.73%	0.82%	0.90%	1.04%
	2SLS		0.01	0.01	0.01	0.01	0.09	0.08	0.07	0.06	11.57%	5.17%	5.97%	6.01%
	Analytical MELO	20	0.01	0.01	0.01	0.01	0.09	0.08	0.07	0.06	11.42%	5.11%	5.93%	5.99%
	Computational MELO		0.01	0.01	0.01	0.01	0.09	0.08	0.07	0.06	11.42%	5.11%	5.94%	5.99%
	2SLS		0.01	0.00	0.00	0.00	0.06	0.05	0.04	0.04	7.62%	3.30%	3.57%	3.81%
	Analytical MELO	50	0.01	0.00	0.00	0.00	0.06	0.05	0.04	0.04	7.58%	3.29%	3.55%	3.80%
	Computational MELO		0.01	0.00	0.00	0.00	0.06	0.05	0.04	0.04	7.57%	3.29%	3.55%	3.80%
	2SLS		0.00	0.00	0.00	0.00	0.04	0.04	0.03	0.03	5.19%	2.34%	2.68%	2.77%
5	Analytical MELO	100	0.00	0.00	0.00	0.00	0.04	0.03	0.03	0.03	5.18%	2.33%	2.67%	2.77%
	Computational MELO		0.00	0.00	0.00	0.00	0.04	0.03	0.03	0.03	5.18%	2.33%	2.67%	2.77%
	2SLS		0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	1.52%	0.73%	0.83%	0.82%
	Analytical MELO	1000	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	1.52%	0.73%	0.83%	0.82%
	Computational MELO		0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	1.52%	0.73%	0.83%	0.82%
	2SLS		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.35%	0.16%	0.18%	0.21%
	Analytical MELO	20000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.35%	0.16%	0.18%	0.21%
	Computational MELO		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.35%	0.16%	0.18%	0.21%

point estimate using the MELO approach, both analytical (Equation 13) and computational (Equation 2 using 10,000 iterations), is 13.14, and with parametric bootstrap (Algorithm A1 using 1,000 as the number of bootstraps), 13.19. The standard deviations are 1.46, and 1.52, respectively, that is, reductions of 22% and 19.5%. These figures are calculated using Equation 10 with Corollary 2.4, and Equation 9, both with Theorem 2.5 in the case of the analytical and computational approaches, and algorithm A1 using parametric bootstrap. Despite the fact that the coefficient of determination in this example is very high ($R^2 = 0.98$), we observe differences between the optimal weight estimates. In addition, the frequentist variability of the optimal weight using the MELO estimates are lower than the one using the *plug-in* approach.

4.2 Space Shuttle Challenger

In 1986, the space shuttle Challenger exploded during take off, killing the seven astronauts aboard. The explosion was the result of an O-ring failure, a splitting of a ring of rubber that seals the parts of the ship together, due to the unusually cold weather (31^oF, i.e., 0^oC) at the time of launch (Dalal et al., 1989).

We calculated the Odds ratio at 45^oF and 69.56^oF (mean sample temperature) for a sample of 23 observations provided by Robert and Casella (2004) taking into account the theoretical structure of subsection 3.2. Using the *plug-in* approach, the probability of failure is 0.996 at 45^oF, therefore the odds ratio estimate is 283.644 with a standard deviation of 999.596. The Odds ratio using MELO is 2.585 in the case of the computational approach (Algorithm A2 setting $B_0 = 10,000\ diag\,\{1,1,1\}$ and $\beta_0 = [0,0,0]$ with 25,000 iterations and a burn-in equal to 5,000), and is 3.886 using parametric bootstrap (Algorithm A1 with 100 as the number of bootstraps and setting $f_{Y_i^*}(\mathbf{y}_i^*|\hat{\beta}_{ML})\mathcal{N}(x_i^T\hat{\beta}_{ML},1)$, see Equations 15 and 16). Observe that the implicit probabilities of the Odds ratio in the MELO approach are 0.721 and 0.795, respectively. However, if the main objective of the statistical inference is the probability, that is, $\mathbf{g}(\theta) = \Phi(x^T\beta)$, which implies $\mathbf{Q}(\theta) = 1$, and $\hat{\omega}^* = E(\Phi(x_i^T\beta))$, we have point estimates equal to 0.964 using the computational MELO, and 0.931 using parametric bootstrap. This highlights a remarkable characteristic of our approach; the estimate depends drastically on the main objective of the inferential situation.

Regarding the frequentist variability of the MELO, we get 2.917 using the computational approach. Observe that in this case, the components associated with Corollary 2.4 depend on the iteration g, so we calculate the mean values over all these components to obtain this figure. We get 2.491 using parametric bootstrap. Observe that there is a huge difference using the delta method (999.596).

The failure probability is 0.266 at 69.56^oF using the *plug-in* approach. This implies an Odds ratio equal to 0.363 with standard deviation equal to 0.171. The Odds ratio using the computational MELO is 0.345 with a standard deviation equal to 0.258, whereas using parametric bootstrap these figures are 0.333 and 0.215, respectively. We get similar point estimates using the central point in the distribution of regressors. In this case, the standard deviation of the *plug-in* is lower than for the MELO approaches (0.33% and 0.204%, respectively).

The message here is that in the case of evaluating a point in the extreme of the distribution of the regressors, that is, when the sample information is not precise (noisy), it is much better to use the MELO approach. On the other hand, it makes sense to use the *plug-in* approach.

4.3 Colonial origins of development

Acemoglu et al. (1993) analyze the effect of property rights on economic growth. They exploit the variability in European settlers' mortality rates during the time of colonization to find the causal effect of protection against expropriation on economic performance. They use 2SLS to accomplish this task. We can write their setting in the following structural system,

$$\begin{aligned} Log(pcGDP)_i =& \beta_0 + \beta_1 PAER_i + \mu_{1i} \\ PAER_i =& \alpha_0 + \alpha_1 Log(pcGDP)_i + \alpha_2 log(Mort)_i + \mu_{2i} \end{aligned}$$

where $pcGDP$, $PAER$ and $Mort$ are the per capita GDP in 1995, the average index of protection against expropriation between 1985 and 1995 (Political Risk Services), and settler mortality rate during the time of colonization (see Acemoglu et al. (1993) for details), respectively. The reduced form model is

$$\begin{aligned} Log(pcGDP)_i =& \pi_0 + \pi_1 log(Mort)_i + e_{1i} \\ PAER_i =& \gamma_0 + \gamma_1 log(Mort)_i + e_{2i} \end{aligned}$$

The first structural equation is exactly identified provided that $\alpha_2 \neq 0$, whereas the second structural equation is sub-identified.

We define the estimation error as $\epsilon = \gamma_1(\hat{\omega} - \beta_1)$, where $\beta_1 = \pi_1/\gamma_1$, then $\mathbf{Q}(\theta) = \gamma_1^2$.

We find the MELO estimates, and their frequentist variability, using the same ideas of subsection 3.4. The outcomes can be seen in Table 7, where we reproduce the outcomes from Acemoglu et al. (1993), Table IV (page 1386), columns 1, 3, 5 and 9.

Table 7: Colonial origins of development

Method	Column (1)[1]	Column (3)[2]	Column (5)[3]	Column (9)[4]
MELO (Computational)[5]	0.91	1.17	0.57	0.95
	(0.14)	(0.27)	(0.09)	(0.15)
MELO (Bootstrap)[6]	0.94	1.21	0.58	0.97
	(0.15)	(0.27)	(0.10)	(0.17)
2SLS	0.94	1.28	0.58	0.98
	(0.16)	(0.36)	(0.10)	(0.17)
Sample size	64	60	37	61
R^2 First stage	0.27	0.13	0.47	0.28

[1] Base sample. [2] Base sample without Neo-Europes. [3] Base sample without Africa.
[4] Base sample, dependent variable is log output per worker. Standard error in parentheses.
[5] Using 10,000 iterations. [6] Using 10,000 iterations and 1,000 bootstraps.

We can see from Table 7 that the standard errors of our approach are always less than the standard errors from 2SLS. We obtain more efficiency gains in noisier models, for instance column (3), where the coefficient of determination is the lowest ($R^2 = 0.13$). In general, the MELO estimates of the effects of property rights on economic performance are lower than the 2SLS estimates.

4.4 Openness and inflation

Romer (1993) analyzes the effect of openness on inflation. In particular, he shows there is a strong and robust negative link between inflation and openness using 2SLS, where he uses the logarithm of the country's land area as an instrument of openness. His model can be written as a structural system of equations:

$$\begin{aligned} Inf_i =& \beta_0 + \beta_1 Open_i + \beta_2 log(pinc_i) + \beta_3 D_i + \mu_{1i} \\ Open_i =& \alpha_0 + \alpha_1 Inf_i + \alpha_2 log(pinc_i) + \alpha_3 log(land_i) + \alpha_4 D_i + \mu_{2i} \end{aligned}$$

where Inf_i, $Open_i$, $pinc_i$, D_i and $land_i$ are the inflation rate, openness, which is measured as the ratio of imports to GDP, real per capita income, and data dummies for the alternative measures of openness and inflation, and land area, respectively.

The first structural equation is exactly identified provided that $\alpha_3 \neq 0$, whereas the second structural equation is sub-identified. The reduced form of this model is

$$\begin{aligned} Inf_i =& \pi_0 + \pi_1 log(pinc_i) + \pi_2 log(land_i) + \pi_3 D_i + e_{1i} \\ Open_i =& \gamma_0 + \gamma_1 log(pinc_i) + \gamma_2 log(land_i) + \gamma_3 D_i + e_{2i} \end{aligned}$$

We define the estimation error as follows:

$$\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} = \begin{bmatrix} \gamma_2(\hat{\omega}_1 - \beta_1) \\ \gamma_2(\hat{\omega}_2 - \beta_2) \end{bmatrix}$$

where $\beta_1 = \pi_2/\gamma_2$, $\beta_2 = \pi_1 - \frac{\pi_2}{\gamma_2}\gamma_1$, then $\mathbf{Q}(\theta) = diag(\gamma_2^2, \gamma_2^2)$.

We can find the MELO estimates, and their frequentist variability, using the same ideas of subsection 3.4. In particular, we have that the structural or causal effect of openness on inflation is -1.252 with a standard error equal to 0.407, and the effect of per capita income on inflation is equal to -0.045 with a standard error of 0.061 (using both the computational and parametric bootstrap with 10,000 iterations, and 1,000 bootstraps). The analogous estimates using 2SLS are -1.260 (0.414) and -0.045 (0.061) for openness and income, respectively.

Despite the fact that MELO and 2SLS are based on completely different frameworks, we practically do not get any differences between these estimates in this application. The reason is that the coefficient of determination in the first stage is equal to 0.48, and there are 100 degrees of freedom. This implies that the signal to noise ratio in the first stage is approximately 1, and given 100 d.f., we are basically replicating the outcomes of our simulation exercise in subsection 3.4 (see Table 6, Signal/Noise=0.5 and Sample size=100).

5 Concluding remarks

Many times the main concern of an econometric inference is associated with non-linear functions of parameters. Our approach tackles directly this issue based on a Bayesian decision theory framework, which allows thinking about the whole inferential situation. Our proposal seems to improve the econometric inference in situations characterized by small sample sizes or noisy models. So, our MELO proposal can be used in situations where getting observations can be a difficult task due to data limitations, for instance, expensive experimental designs or availability restrictions, and/or situations where the models are very noisy, for instance, very weak instruments. But, if there is a moderate sample size and/or the models are very informative, it is better to use the commonly used alternatives, due to the availability of the appropriate software for them.

However, we must acknowledge that our approach requires an analytical solution for $\mathbf{g}(\theta)$. This is problematic in some complex settings. In addition, the non-existence of sufficient statistics necessitates an extra computational burden in order to apply parametric bootstrap. Future research should find general asymptotic results for our MELO proposal as well as the development of a theoretical framework for over-identified models.

References

Acemoglu, D., Johnson, S., and Robinson, J. (1993). The colonial origins of comparative development: An empirical investigation. *The American Economic Review*, 91(5):1369–1401.

Bayarri, M. J. and Berger, J. (2004). The interplay of Bayesian and frequentist analysis. *Statistical Science*, 19(1):58–80.

Berger, J. (1993). *Statistical Decision Theory and Bayesian Analysis.* third edition. Springer-Verlag, Berlin.

Bernardo, J. and Smith, A. (1994). *Bayesian Theory.* Wiley, Chichester, UK.

Claeskens, G. and Hjort, N. L. (2003). The focused information criterion. *Journal of the American Statistical Association*, 98(464):900–916.

Dalal, S. R., Fowlkes, E. B., and Hoadley, B. (1989). Risk analysis of the space shuttle: Pre-Challenger prediction of failure. *Journal of the American Statistical Association*, 84(408):945–957.

Diebold, F. X. and Lamb, R. L. (1997). Why are estimates of agricultural supply response so variable? *Journal of Econometrics*, 76(1):357–373.

DiTraglia, F. J. (2016). Using invalid instruments on purpose: Focused moment selection and averaging for GMM. *Journal of Econometrics*, 195(2):187–208.

Drèze, J. (1974). Econometrics and decision theory. In Fienberg, S. E. and Zellner, A., editors, *Studies in Bayesian Econometrics and Statistics.* North-Holland, Amsterdam. pp. 17–38.

Efron, B. (2012). Bayesian inference and parametric bootstrap. *The Annals of Applied Statistics*, 6(4):1971–1997.

Efron, B. (2015). Frequentist accuracy of Bayesian estimates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(3):617–646.

Good, I. J. (1992). The Bayes/non Bayes compromise: A brief review. *Journal of the American Statistical Association*, 87(419):597–606.

Greenberg, E. (2012). *Introduction to Bayesian Econometrics.* Cambridge University Press.

Hansen, B. E. (2005). Challenges for econometric model selection. *Econometric Theory*, 21(01):60–68.

Judge, G., Carter, R., Griffiths, W., Lutkeploh, H., and Lee, T.-C. (1988). *Introduction to the Theory and Practice of Econometrics.* second edition. Wiley, New York.

Kass, R. (2011). Statistical inference: the big picture. *Statistical Science*, 26(1):1–9.

Marschak, J. (1960). Economic measurements for policy and prediction. In Holt, C. C., Modigliani, F., Muth, J. F., and Simon, H. A., editors, *Planning Production, Inventories and Workforce*. Prentice-Hall, Englewood Cliffs, NJ. pp. 1–26.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Linear generalized models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3):370–384.

Park, S. B. (1982). Some sampling properties of minimum expected loss (MELO) estimators of structural coefficients. *Journal of Econometrics*, 18(3):295–311.

Ramírez Hassan, A. (2017). The interplay between the Bayesian and frequentist approaches: A general nesting spatial panel data model. *Spatial Economic Analysis*, 12(1):92–112.

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Second edition. Springer-Verlag, New York.

Romer, D. (1993). Openness and inflation: Theory and evidence. *The Quarterly Journal of Economics*, 108(4):869–903.

Swamy, P. A. V. B. and Mehta, J. S. (1983). Further results on Zellner's minimum expected loss and full information maximum likelihood estimators for undersized samples. *Journal of Business & Economic Statistics*, 1(2):154–162.

Uriel, E., Contreras, D., Moltó, M., and Peiró, A. (1997). *Econometría El modelo lineal*. third edition. Editorial AC, Madrid.

Wald, A. (1945). Generalization of a theorem by v.Neumann concerning zero sum two-person games. *Annals of Mathematics*, 46(2):281–286.

Wald, A. (1947). Foundations of a general theory of sequential decision functions. *Econometrica*, 15(4):279–313.

Zellner, A. (1978). Estimation of functions of population means and regression coefficients including structural coefficients. *Journal of econometrics*, 8:127–158.

Zellner, A. (1996). *An Introduction to Bayesian Inference in Econometrics*. Wiley, New York.

Zellner, A. (1998). The finite sample properties of simultaneous equations' estimates and estimators Bayesian and non-Bayesian approaches. *Journal of Econometrics*, 83(1):185–212.

Zellner, A. and Park, S. B. (1979). Minimum expected loss (MELO) estimators for functions of parameters and structural coefficients of econometric models. *Journal of the American Statistical Association*, 74(365):185–193.

Zellner, A. and Park, S. B. (1980). A note on the relationship of minimum expected loss (MELO) and other structural coefficient estimates. *The Review of Economics and Statistics*, 62(3):482–484.

6 Appendix

6.1 Proof of Theorem 2.1

Given $\mathcal{L}(\mathbf{g}(\theta), \hat{\omega}) = (\mathbf{g}(\theta) - \hat{\omega})^T \mathbf{Q}(\theta)(\mathbf{g}(\theta) - \hat{\omega})$, the posterior expected value of the loss function is

$$E_{\pi_\theta(\theta|\mathbf{y})}\{\mathcal{L}(\mathbf{g}(\theta), \hat{\omega})\} = E_{\pi_\theta(\theta|\mathbf{y})}\left\{\mathbf{g}(\theta)^T \mathbf{Q}(\theta)\mathbf{g}(\theta)\right\} - 2\hat{\omega}^T E_{\pi_\theta(\theta|\mathbf{y})}\{\mathbf{Q}(\theta)\mathbf{g}(\theta)\} + \hat{\omega}^T E_{\pi_\theta(\theta|\mathbf{y})}\{\mathbf{Q}(\theta)\}\hat{\omega}$$

then

$$\frac{\partial E_{\pi_\theta(\theta|\mathbf{y})}\{\mathcal{L}(\mathbf{g}(\theta), \hat{\omega})\}}{\partial \hat{\omega}} = -2E_{\pi_\theta(\theta|\mathbf{y})}\{\mathbf{Q}(\theta)\mathbf{g}(\theta)\} + 2E_{\pi_\theta(\theta|\mathbf{y})}\{\mathbf{Q}(\theta)\}\hat{\omega}^* = \mathbf{0}$$

so,

$$\hat{\omega}^*(\mathbf{y}) = \left[E_{\pi_\theta(\theta|\mathbf{y})}\mathbf{Q}(\theta)\right]^{-1} E_{\pi_\theta(\theta|\mathbf{y})}[\mathbf{Q}(\theta)\mathbf{g}(\theta)]$$

observe that

$$\frac{\partial^2 E_{\pi_\theta(\theta|\mathbf{y})}\{\mathcal{L}(\mathbf{g}(\theta), \hat{\omega})\}}{\partial \hat{\omega}\partial \hat{\omega}^T} = 2E_{\pi_\theta(\theta|\mathbf{y})}\{\mathbf{Q}(\theta)\}$$

6.2 Proof of Theorem 2.2

$$\hat{\omega}^*(\mathbf{y}) = \left[\int_\Theta \mathbf{Q}(\theta)\pi_\theta(\theta|\mathbf{y})d\theta\right]^{-1}\left[\int_\Theta \mathbf{Q}(\theta)\mathbf{g}(\theta)\pi_\theta(\theta|\mathbf{y})d\theta\right]$$

where

$$\begin{aligned}\pi_\theta(\theta|\mathbf{y}) &= \frac{\pi(\theta)f(\mathbf{y}|\theta)}{\int \pi(\theta)f(\mathbf{y}|\theta)d\theta} \\ &= \frac{\pi(\theta)w(\mathbf{y})h(\hat{\theta}|\theta)}{\int \pi(\theta)w(\mathbf{y})h(\hat{\theta}|\theta)d\theta}\end{aligned}$$

The second equality follows from the Factorization theorem due to $\hat{\theta}$ is a sufficient statistic, $w(\mathbf{y})$ does not depend on θ, and $h(\hat{\theta}|\theta)$ depends on $\mathbf{y}$ only through $\hat{\theta}$. Then,

$$\begin{aligned}\hat{\omega}^*(\mathbf{y}) &= \left[\int_\Theta \mathbf{Q}(\theta)\frac{\pi(\theta)w(\mathbf{y})h(\hat{\theta}|\theta)}{\int \pi(\theta)w(\mathbf{y})h(\hat{\theta}|\theta)d\theta}d\theta\right]^{-1}\left[\int_\Theta \mathbf{Q}(\theta)\mathbf{g}(\theta)\frac{\pi(\theta)w(\mathbf{y})h(\hat{\theta}|\theta)}{\int \pi(\theta)w(\mathbf{y})h(\hat{\theta}|\theta)d\theta}d\theta\right] \\ &= \left[\int_\Theta \mathbf{Q}(\theta)\pi(\theta)h(\hat{\theta}|\theta)d\theta\right]^{-1}\left[\int_\Theta \mathbf{Q}(\theta)\mathbf{g}(\theta)\pi(\theta)h(\hat{\theta}|\theta)d\theta\right] \\ &= \left[\int_\Theta \mathbf{Q}(\theta)\pi_\theta(\theta|\hat{\theta})d\theta\right]^{-1}\left[\int_\Theta \mathbf{Q}(\theta)\mathbf{g}(\theta)\pi_\theta(\theta|\hat{\theta})d\theta\right] \\ &= \left[E_{\pi_\theta(\theta|\hat{\theta}(\mathbf{y}))}\mathbf{Q}(\theta)\right]^{-1} E_{\pi_\theta(\theta|\hat{\theta}(\mathbf{y}))}[\mathbf{Q}(\theta)\mathbf{g}(\theta)] \\ &= \hat{\omega}^*(\hat{\theta}(\mathbf{y}))\end{aligned}$$

The third equality is due to the likelihood principle, that is, all the relevant information regarding θ is in $h(\hat{\theta}|\theta)$ (Berger, 1993; Bernardo and Smith, 1994).

6.3 Proof of Lemma 2.3

$$
\begin{aligned}
\hat{\omega}^*(\mathbf{y}) = \hat{\omega}^*(\hat{\theta}(\mathbf{y})) &= [E_{\pi_\theta(\theta|\hat{\theta}(\mathbf{y}))}\mathbf{Q}(\theta)]^{-1}E_{\pi_\theta(\theta|\hat{\theta}(\mathbf{y}))}[\mathbf{Q}(\theta)\mathbf{g}(\theta)] \\
&= \left\{\int \mathbf{Q}(\theta)\pi(\theta)h(\hat{\theta}|\theta)d\theta\right\}^{-1}\left\{\int \mathbf{Q}(\theta)\mathbf{g}(\theta)\pi(\theta)h(\hat{\theta}|\theta)d\theta\right\}
\end{aligned}
$$

Denoting

$$
\begin{aligned}
A &= \int \mathbf{Q}(\theta)\pi(\theta)h(\hat{\theta}|\theta)d\theta \\
B &= \int \mathbf{Q}(\theta)\mathbf{g}(\theta)\pi(\theta)h(\hat{\theta}|\theta)d\theta
\end{aligned}
$$

of order $K \times K$ and $K \times 1$ respectively. So, $\hat{\omega} = A^{-1}B$, applying the properties of the differentiation of matrices,[2]

$$
\nabla_{\hat{\theta}(\mathbf{y})}\hat{\omega}^*(\hat{\theta}(\mathbf{y})) = A^{-1}\nabla_{\hat{\theta}(\mathbf{y})}B + \nabla_{\hat{\theta}(\mathbf{y})}A^{-1}\left[B \otimes I_P\right]
$$

where

$$
\nabla_{\hat{\theta}(\mathbf{y})}A^{-1} = -A^{-1}\nabla_{\hat{\theta}(\mathbf{y})}A\left[A^{-1} \otimes I_N\right]\ ^{3}
$$

Therefore,[4]

$$
\begin{aligned}
\nabla_{\hat{\theta}(\mathbf{y})}\hat{\omega}^*(\hat{\theta}(\mathbf{y})) &= A^{-1}\nabla_{\hat{\theta}(\mathbf{y})}B - A^{-1}\nabla_{\hat{\theta}(\mathbf{y})}A\left[A^{-1} \otimes I_P\right]\left[B \otimes I_P\right] \\
&= A^{-1}\left\{\nabla_{\hat{\theta}(\mathbf{y})}B - \nabla_{\hat{\theta}(\mathbf{y})}A\left[A^{-1} \otimes I_P\right]\left[B \otimes I_P\right]\right\} \\
&= A^{-1}\left\{\nabla_{\hat{\theta}(\mathbf{y})}B - \nabla_{\hat{\theta}(\mathbf{y})}A\left[A^{-1}B \otimes I_P\right]\right\} \\
&= A^{-1}\left\{\nabla_{\hat{\theta}(\mathbf{y})}B - \nabla_{\hat{\theta}(\mathbf{y})}A\left[\hat{\omega} \otimes I_P\right]\right\}
\end{aligned}
$$

where $\nabla_{\hat{\theta}(\mathbf{y})}\hat{\omega}^*(\mathbf{y})$ is a $K \times P$ matrix of partial derivatives, that is,

$$
\nabla_{\hat{\theta}(\mathbf{y})}\hat{\omega}^*(\hat{\theta}(\mathbf{y})) = \begin{bmatrix}
\frac{\partial\hat{\omega}_1^*}{\partial\hat{\theta}(\mathbf{y})_1} & \frac{\partial\hat{\omega}_1^*}{\partial\hat{\theta}(\mathbf{y})_2} & \cdots & \frac{\partial\hat{\omega}_1^*}{\partial\hat{\theta}(\mathbf{y})_P} \\
\frac{\partial\hat{\omega}_2^*}{\partial\hat{\theta}(\mathbf{y})_1} & \frac{\partial\hat{\omega}_2^*}{\partial\hat{\theta}(\mathbf{y})_2} & \cdots & \frac{\partial\hat{\omega}_2^*}{\partial\hat{\theta}(\mathbf{y})_P} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial\hat{\omega}_K^*}{\partial\hat{\theta}(\mathbf{y})_1} & \frac{\partial\hat{\omega}_K^*}{\partial\hat{\theta}(\mathbf{y})_2} & \cdots & \frac{\partial\hat{\omega}_K^*}{\partial\hat{\theta}(\mathbf{y})_P}
\end{bmatrix},
$$

$$
\begin{aligned}
\nabla_{\hat{\theta}(\mathbf{y})}A &= \int \mathbf{Q}(\theta)\pi(\theta)\nabla_{\hat{\theta}(\mathbf{y})}h(\hat{\theta}|\theta)d\theta \\
&= \int (\mathbf{Q}(\theta) \otimes \alpha_{\hat{\theta}(\mathbf{y})}(\theta))\pi(\theta)h(\hat{\theta}|\theta)d\theta
\end{aligned}
$$

[2] If A and B are matrices of order $N \times K$ and $K \times M$, and x is a vector of order $1 \times N$, then $\nabla_x AB = \nabla_x A[B \otimes I_N] + A\nabla_x B$.

[3] $0 = \nabla_{\hat{\theta}(\mathbf{y})}AA^{-1} = \nabla_{\hat{\theta}(\mathbf{y})}A[A^{-1} \otimes I_P] + A^{-1}\nabla_{\hat{\theta}(\mathbf{y})}A^{-1}$, then $\nabla A^{-1} = -A^{-1}\nabla_{\hat{\theta}(\mathbf{y})}A\left[A^{-1} \otimes I_P\right]$.

[4] For the outcome of the third equality, let A_1, A_2, B_1 and B_2 be matrices of orders $M \times N$, $M \times P$, $L \times R$ and $R \times P$, then $(A_1 \otimes B_1)(A_2 \otimes B_2) = A_1A_2 \otimes B_1B_2$.

and

$$
\begin{aligned}
\nabla_{\hat{\theta}(\mathbf{y})} B &= \int \mathbf{Q}(\theta)\mathbf{g}(\theta)\pi(\theta)\nabla_{\hat{\theta}(\mathbf{y})} h(\hat{\theta}|\theta)d\theta \\
&= \int ((\mathbf{Q}(\theta)\mathbf{g}(\theta)) \otimes \alpha_{\hat{\theta}(\mathbf{y})}(\theta))\pi(\theta)h(\hat{\theta}|\theta)d\theta
\end{aligned}
$$

so,

$$
\begin{aligned}
\nabla_{\hat{\theta}(\mathbf{y})}\hat{\omega}^*(\hat{\theta}(\mathbf{y})) &= \left\{\int \mathbf{Q}(\theta)\pi(\theta)h(\hat{\theta}|\theta)d\theta\right\}^{-1} \\
&\times \left\{\int ((\mathbf{Q}(\theta)\mathbf{g}(\theta)) \otimes \alpha_{\hat{\theta}(\mathbf{y})}(\theta))\pi(\theta)h(\hat{\theta}|\theta)d\theta - \left[\int (\mathbf{Q}(\theta) \otimes \alpha_{\hat{\theta}(\mathbf{y})}(\theta))\pi(\theta)h(\hat{\theta}|\theta)d\theta\right][\hat{\omega} \otimes I_P]\right\} \\
&= \left\{E_{\pi_\theta(\hat{\theta}(\mathbf{y}))}[Q(\theta)]\right\}^{-1}\left\{E_{\pi_\theta(\hat{\theta}(\mathbf{y}))}[(Q(\theta)g(\theta)) \otimes \alpha_{\hat{\theta}(\mathbf{y})}(\theta)] - E_{\pi_\theta(\hat{\theta}(\mathbf{y}))}[Q(\theta) \otimes \alpha_{\hat{\theta}(\mathbf{y})}(\theta)][\hat{\omega} \otimes I_P]\right\}
\end{aligned}
$$

where

$$
E[(Q(\theta)g(\theta)) \otimes \alpha_{\hat{\theta}(\mathbf{y})}(\theta)] =
$$

$$
\begin{bmatrix}
E\sum_{j=1}^K Q_{1j}g_j \frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_1} & E\sum_{j=1}^K Q_{1j}g_j \frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_2} & \dots & E\sum_{j=1}^K Q_{1j}g_j \frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_P} \\
E\sum_{j=1}^K Q_{2j}g_j \frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_1} & E\sum_{j=1}^K Q_{2j}g_j \frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_2} & \dots & E\sum_{j=1}^K Q_{2j}g_j \frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_P} \\
\vdots & \vdots & \ddots & \vdots \\
E\sum_{j=1}^K Q_{Kj}g_j \frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_1} & E\sum_{j=1}^K Q_{Kj}g_j \frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_2} & \dots & E\sum_{j=1}^K Q_{Kj}g_j \frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_P}
\end{bmatrix}
$$

and

$$
E[Q(\theta) \otimes \alpha_{\hat{\theta}(\mathbf{y})}(\theta)] =
$$

$$
\begin{bmatrix}
EQ_{11}\frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_1} & EQ_{11}\frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_2} & \dots & EQ_{11}\frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_P} & EQ_{12}\frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_1} & EQ_{12}\frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_2} & \dots & EQ_{1K}\frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_P} \\
EQ_{21}\frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_1} & EQ_{21}\frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_2} & \dots & EQ_{21}\frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_P} & EQ_{22}\frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_1} & EQ_{22}\frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_2} & \dots & EQ_{2K}\frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_P} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
EQ_{K1}\frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_1} & EQ_{K1}\frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_2} & \dots & EQ_{K1}\frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_P} & EQ_{K2}\frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_1} & EQ_{K2}\frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_2} & \dots & EQ_{KK}\frac{\partial logh(\hat{\theta}|\theta)}{\partial\hat{\theta}(\mathbf{y})_P}
\end{bmatrix}
$$

6.4 Proof Corollary 2.4

$$
\begin{aligned}
\hat{\omega}^*(\hat{\theta}(\mathbf{y})) &= \frac{E_{\pi_\theta(\hat{\theta}(\mathbf{y}))}[Q(\theta)g(\theta)|\hat{\theta}(\mathbf{y})]}{E_{\pi_\theta(\hat{\theta}(\mathbf{y}))}[Q(\theta)|\hat{\theta}(\mathbf{y})]} \\
&= \frac{\left[\int Q(\theta)g(\theta)\pi(\theta)h(\hat{\theta}|\theta)d\theta\right]/c}{\left[\int Q(\theta)\pi(\theta)h(\hat{\theta}|\theta)d\theta\right]/c}
\end{aligned}
$$

where

$$c = \int \pi(\theta) h(\hat{\theta}|\theta) d\theta$$

.

Let

$$\begin{aligned} A &= \int Q(\theta) g(\theta) \pi(\theta) h(\hat{\theta}|\theta) d\theta \\ B &= \int Q(\theta) \pi(\theta) h(\hat{\theta}|\theta) d\theta. \end{aligned}$$

then

$$\begin{aligned} A' &= \int Q(\theta) g(\theta) \alpha_{\hat{\theta}(\mathbf{y})}(\theta) \pi(\theta) h(\hat{\theta}|\theta) d\theta \\ B' &= \int Q(\theta) \alpha_{\hat{\theta}(\mathbf{y})}(\theta) \pi(\theta) h(\hat{\theta}|\theta) d\theta \end{aligned}$$

Therefore,

$$\begin{aligned} \hat{\omega}^*(\hat{\theta}(\mathbf{y}))' =& \frac{A}{B} \left\{ \frac{A'}{A} - \frac{B'}{B} \right\} \\ =& \frac{E_{\pi_\theta(\theta|\hat{\theta}(\mathbf{y}))}[Q(\theta) g(\theta)|\hat{\theta}(\mathbf{y})]}{E_{\pi_\theta(\theta|\hat{\theta}(\mathbf{y}))}[Q(\theta)|\hat{\theta}(\mathbf{y})]} \\ & \times \left\{ \frac{E_{\pi_\theta(\theta|\hat{\theta}(\mathbf{y}))}[Q(\theta) g(\theta) \alpha_{\hat{\theta}(\mathbf{y})}(\theta)|\hat{\theta}(\mathbf{y})]}{E_{\pi_\theta(\theta|\hat{\theta}(\mathbf{y}))}[Q(\theta) g(\theta)|\hat{\theta}(\mathbf{y})]} - \frac{E_{\pi_\theta(\theta|\hat{\theta}(\mathbf{y}))}[Q(\theta) \alpha_{\hat{\theta}(\mathbf{y})}(\theta)|\hat{\theta}(\mathbf{y})]}{E_{\pi_\theta(\theta|\hat{\theta}(\mathbf{y}))}[Q(\theta)|\hat{\theta}(\mathbf{y})]} \right\} \\ =& \frac{E_{\pi_\theta(\theta|\hat{\theta}(\mathbf{y}))}[Q(\theta) g(\theta) \alpha_{\hat{\theta}(\mathbf{y})}(\theta)|\hat{\theta}(\mathbf{y})]}{E_{\pi_\theta(\theta|\hat{\theta}(\mathbf{y}))}[Q(\theta)|\hat{\theta}(\mathbf{y})]} \\ & - \frac{E_{\pi_\theta(\theta|\hat{\theta}(\mathbf{y}))}[Q(\theta) g(\theta)|\hat{\theta}(\mathbf{y})] E_{\pi_\theta(\theta|\hat{\theta}(\mathbf{y}))}[Q(\theta) \alpha_{\hat{\theta}(\mathbf{y})}(\theta)|\hat{\theta}(\mathbf{y})]}{(E_{\pi_\theta(\theta|\hat{\theta}(\mathbf{y}))}[Q(\theta)|\hat{\theta}(\mathbf{y})])^2} \end{aligned}$$

6.5 Proof of Theorem 2.5

Given a sufficient statistic such that $\hat{\theta}(\mathbf{y}) \sim (\mu_\theta, \Sigma_\theta)$, then a second order Taylor expansion implies that

$$\hat{\omega}^*(\mathbf{y}) = \hat{\omega}^*(\hat{\theta}(\mathbf{y})) \approx \hat{\omega}^*(\theta_0) + \nabla_\theta \hat{\omega}^*(\theta_0)(\hat{\theta} - \theta_0)$$

Then the variance of $\hat{\omega}^*(\hat{\theta}(\mathbf{y}))$ is

$$Var(\hat{\omega}^*(\mathbf{y})) = Var(\hat{\omega}^*(\hat{\theta}(\mathbf{y}))) \approx \nabla_\theta \hat{\omega}^*(\theta_0) \Sigma_{\hat{\theta}} \nabla_\theta \hat{\omega}^*(\theta_0)^T$$

if $\hat{\theta} \xrightarrow{p} \theta_0$, then

$$Var(\hat{\omega}^*(\mathbf{y})) = Var(\hat{\omega}^*(\theta(\mathbf{y}))) \approx \nabla_\theta \hat{\omega}^*(\hat{\theta}) \Sigma_{\hat{\theta}} \nabla_\theta \hat{\omega}^*(\hat{\theta})^T$$