Centre for Efficiency and Productivity Analysis

Working Paper Series
No. WP01/2022

Testing for Optimization Behavior in Production when Data is with Measurement Errors: A Bayesian Approach

Mike G. Tsionas and Valentin Zelenyuk

Date: January 2022

School of Economics
University of Queensland
St. Lucia, Qld. 4072
Australia

ISSN No. 1932 - 4398

Testing for Optimization Behavior in Production when Data is with Measurement Errors: A Bayesian Approach

Mike G. Tsionas* Valentin Zelenyuk†

January 18, 2022

Abstract

The purpose of this paper is to develop formal tests for cost / profit rationalization of observed data sets under measurement errors in both prices and quantities. The new techniques are based on new statistical formulations for inequalities that describe cost and profit rationalizability, developed in a Bayesian framework. The new likelihood-based methods of inference are introduced and then illustrated using a data set of large U.S. banks. We also develop various robustness checks, including a normal and lognormal specification of the data generating process, as well as a multivariate mixture-of-normal-distributions.

Key Words: Cost Minimization; Profit Maximization; Likelihood-based methods; Markov Chain Monte Carlo; Banking.

*Montpellier Business School & Lancaster University Management School, LA1 4YX, U.K., m.tsionas@lancaster.ac.uk

†Valentin Zelenyuk, School of Economics, The University of Queensland, Brisbane, Australia; v.zelenyuk@uq.edu.au.
The support from the Australian Research Council (FT170100401) and The University of Queensland is acknowledged. Feedback from Bao Hoang Nguyen, Evelyn Smart and Zhichao Wang is appreciated. These individuals and organizations are not responsible for the views expressed here.

1 Introduction

Testing a data set for compatibility with cost minimization or profit maximization is essential in many areas of economics and operations research. Measuring or estimating the departures from cost minimization or profit maximization is also an essential part of theoretical and applied studies and a large amount of literature has accumulated over the years that provides methods for estimating departures from the standard behavioral assumptions.[1]

Quite often attempts to consolidate observed data with the standard behavioral assumptions face the problem of measurement error in both quantities and prices yielding highly complicated mathematical programming problems whose statistical properties are also largely unknown (e.g. Varian, 1982a,b,c, 1983a,b, 1984, 1985).

As one of the prominent examples, Varian (1985) proposed a test for whether a data set of input prices $w_i \in \mathbb{R}_+^K$, input decisions $x_i \in \mathbb{R}_+^K$ and produced outputs $y_i \in \mathbb{R}_+^M$ $(i = 1, \ldots, n)$ can be rationalized by cost minimization (*c-rationalizable*). Also, recall that Varian's (1984) Weak Axiom of Cost Minimization (WACM) is a necessary and sufficient condition for the observed behavior of the firm to be compatible with cost minimizing behavior:

$$w_i' x_i \leq w_i' x_j \text{ for all } i, j \text{ such that } y_i \leqq y_j. \tag{1}$$

For future reference, we define $\mathcal{A} = \{(i, j) \in \{1, \ldots, n\} | y_i \leqq y_j\}$ with cardinality N, after excluding cases where we compare the same output to itself, and $\mathcal{A}_i = \{j \in \{1, \ldots, n\} | y_i \leqq y_j\}$ with cardinality N_j. Varian also assumed that input quantities are measured with error so that, for example,

$$x_{ik} = \chi_{ik} + e_{ik},\ i = 1, \ldots, n,\ k = 1, \ldots, K, \tag{2}$$

[1] E.g., see Afriat (1967, 1972), Hanoch and Rothschild (1972), Varian (1982a,b,c, 1983a,b, 1984, 1985), Chavas and Cox (1990), Färe and Grosskopf (1985), to mention a few.

where $\chi_i \in \mathbb{R}_+^K$ are true but unobservable input decisions and e_{ik} represents measurement error, distributed as $e_{ik} \sim$ i.i.d. $\mathcal{N}(0, \sigma_e^2)$. In turn, Varian proposed a clever mathematical programming problem to obtain a test statistic for c-rationalizability, given by:

$$\begin{aligned} &\min_{\{\chi_{ik}\}} \tfrac{1}{\sigma_e^2} \textstyle\sum_{i=1}^n \left(x_i - \chi_i\right)^2, \\ &\text{s.t. } w_i' \chi_i \leq w_i' \chi_j \text{ for all } i, j \text{ such that } y_i \leqq y_j. \end{aligned} \tag{3}$$

This is a quadratic programming problem in which the test statistic depends on the critical parameter σ_e^2 which is, however, unknown. Varian responds to this point as follows:

> "Does anyone really believe that the factor demand data described in table 1 were measured with a standard error of less than 0.05 percent? If not, the procedures outlined above indicate that the departures from cost minimizing behavior depicted in that table are not statistically significant. This seems to me to be a perfectly satisfactory statement. Furthermore, specifying the likely magnitude of the measurement error seems to me to be much *less difficult* a task than specifying a plausible functional form for a production function, as is required in the conventional approach." (Varian, 1985, p. 456).

To the criticism that he assumes the presence of measurement error in factor decisions but not in input prices, his reply was:

> "I agree. It would be desirable to incorporate error terms on the prices as well. However, note that the resulting programming problem would then have non-linear constraints and thus be considerably more difficult to solve. Furthermore, note that standard regression methods typically specify that regressors are non-stochastic. If one is estimating conditional factor demand equations, this means that price and output variables are hypothesized to be measured without error, and only factor demands themselves are assumed to be measured with error - exactly as specified here." (Varian, 1985, p. 456).

Although the first point can be thought of as “fair enough”, the second is more disputable these days as, typically, economists tend to look into endogeneity and related issues much more closely than in the past. Varian’s (1984) technique can be extended to profit maximization. So, we say that we have p-rationalizability (Weak Axiom of Profit Maximization, WAPM) if and only if

$$\mathbf{p}_i'\mathbf{y}_i \geq \mathbf{p}_i'\mathbf{y},\ \forall i \in \{1, \ldots, n\}\,,\ \forall \mathbf{y} \in \boldsymbol{Y}, \tag{4}$$

where $\mathbf{p}_i$ is a vector that contains input and output prices corresponding to netputs in vector $\mathbf{y}_i$ (Varian, 1984, Theorem 3, p. 584) and the technology set $\boldsymbol{Y}$ is defined as: $\mathbf{y} \in \boldsymbol{Y} \Rightarrow \widetilde{\mathbf{y}} \in \boldsymbol{Y}$, $\widetilde{\mathbf{y}} \leqq \mathbf{y}$, i.e., only free disposability of inputs and outputs is imposed. Therefore, (4) holds for any netput bundle that is technically feasible when $\mathbf{y}_i$ was chosen. We will assume this throughout without mentioning it to avoid repetition.

Varian’s work was successfully utilized in several contributions (Keshvari, 2017, Keshvari and Kuosmanen, 2013, Kuosmanen, 2008, Kuosmanen and Kortelainen, 2012; see also Yagi et al. (2020), Lee et al., 2013, Du et al., 2013, Smeulders et al., 2019, etc.) Meanwhile, Färe and Grosskopf (1995) showed more explicitly the analogy between Varian’s approach and the activity analysis modeling for efficiency analysis. More recently, Echenique et al. (2019) introduced the idea of a “money pump” to argue that many violations of WARP indeed occur, however, they are small in monetary terms.

In this work we extend the c-rationalizability and p-rationalizability by exploiting an explicit data generating process for the data. Then, we incorporate the weak axioms of cost minimization or profit maximization as priors on latent, unobserved prices and quantities. Finally, we use Bayesian analysis based on Markov Chain Monte Carlo (MCMC) to derive exact marginal posterior distributions for compatibility with c-rationalizability or p-rationalizability and measure the extent of departures from these behavioral assumptions.

2 The Model

2.1 General ideas

Generalizing Varian (1985), we assume that both prices and quantities are measured with error:

$$\begin{bmatrix} x_i \\ w_i \end{bmatrix} = \begin{bmatrix} \chi_i \\ \omega_i \end{bmatrix} + \varepsilon_i,\ \ \varepsilon_i \sim \text{i.i.d. } \mathcal{N}_{2K}\left(\mathbf{0}, \Sigma)\right), \tag{5}$$

where ε_i represents a vector of unknown idiosyncratic noise or measurement error, supported in $\mathbb{R}^{2K}$, where we allow for a general covariance matrix Σ (as it is difficult to argue that input decisions and input prices are uncorrelated), and χ_i, ω_i represents the true but unobservable values of input decisions and input prices, respectively. For the unobservable quantities and prices we assume

$$\begin{bmatrix} \chi_i \\ \omega_i \end{bmatrix} \sim \text{i.i.d. } \mathcal{N}\left(\mu, \Omega\right),\ i = 1, \ldots, n, \tag{6}$$

where $\mu \in \mathbb{R}^{2K}$ represents the mean vector and Ω is the covariance matrix. Moreover, $\mu = \begin{bmatrix} \mu_\chi \\ \mu_\omega \end{bmatrix}$.[2] So, the unobserved values of input decisions and input prices are allowed to be correlated. We can partition Σ and Ω in the obvious way:

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{bmatrix},\ \Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}' & \Omega_{22} \end{bmatrix}, \tag{7}$$

[2] Although normality assumptions may seem strong, in fact, they can be easily generalized by using mixtures of multivariate normal distributions leading, effectively, to non-parametric models (Geweke and Keane, 2007; Norets, 2012).

and the mean vector, $\mu = [\mu_\chi', \mu_\omega']'$. As we treat χ_is and ω_is as parameters it seems natural to place prior information for them by exploiting the WACM in the constraints[3] of (3). The prior information we impose is inspired by the stochastic frontier literature[4] and is as follows:

$$\omega_i'\chi_i = \omega_i'\chi_j + v_{ij} - u_{ij},\ \forall (i,j) \in \mathcal{A}, \tag{8}$$

where $v_{ij} \sim$ i.i.d.$\mathcal{N}(0, \sigma_v^2)$, and $u_{ij} \sim$ i.i.d.$\mathcal{N}_+(0, \sigma_u^2)$, for all $(i,j) \in \mathcal{A}$. Moreover, we define

$$\lambda = \tfrac{\sigma_u}{\sigma_v}. \tag{9}$$

So, our prior states that there are likely to be two-sided error deviations from WACM captured by v_{ij} but there is also a non-negative error component u_{ij} which ensures that WACM holds. The assumption of half-normality is not as strong as it seems (Echenique et al., 2011, p. 1211) although we remove it in Section 5. Moreover, according to Fleissig and Whitney (2003) few violations of the Generalized Axiom of Revealed Preference are due to measurement error even when this error is substantial.

[3] In the linear model $y = X\beta + \varepsilon$ in obvious notation, one's prior may be the single relationship that a linear combination $c'\beta \sim \mathcal{N}(0, \sigma_o^2)$ for some vector c conformable with β.

[4] See Aigner, Lovell and Schmidt (1977), Cornwell, Schmidt and Sickles (1990) and a comparison of parametric, semi- and non-parametric SFA approaches in Parmeter and Zelenyuk (2019), to mention a few.

Under these assumptions we can formulate a posterior which is as follows:

$$
\begin{gathered}
p(\theta, \chi, \omega, u; \mathcal{D}) \propto \\
|\Sigma|^{-n/2} \exp\left\{ -\frac{1}{2}\sum_{i=1}^{n} \left[(x_i - \chi_i)' \quad (w_i - \omega_i)' \right] \Sigma^{-1} \left[\begin{array}{c} x_i - \chi_i \\ \\ w_i - \omega_i \end{array} \right] \right\} \cdot \\
\sigma_v^{-N} \exp\left\{ -\frac{1}{2\sigma_v^2} \sum_{(i,j)\in\mathcal{A}} \left(\omega_i' \chi_i - \omega_i' \chi_j + u_{ij} \right) \right\} \cdot \\
\sigma_u^{-N} \exp\left\{ -\frac{1}{2\sigma_u^2} \sum_{(i,j)\in\mathcal{A}} u_{ij}^2 \right\} \cdot \\
|\Omega|^{-n/2} \exp\left\{ -\frac{1}{2}\sum_{i=1}^{n} \left[(x_i - \mu_\chi)' \ (w_i - \mu_\omega)' \right] \Omega^{-1} \left[\begin{array}{c} x_i - \mu_\chi \\ \\ w_i - \mu_\omega \end{array} \right] \right\} \cdot p(\theta),
\end{gathered} \tag{10}
$$

where $\theta = (\sigma_v, \sigma_u, \Sigma, \Omega, \mu)$ represents the set of our "structural parameters" for which we define a prior $p(\theta)$, $\mathcal{D} = \{(x_i, w_i)\}_{i=1}^{n}$ represents the available data, $\chi = [\chi_1', \ldots, \chi_n']'$, $\omega = [\omega_1', \ldots, \omega_n']'$, and $u = [u_{ij}, (i,j) \in \mathcal{A}]$. It is worth noting here that the constraint $y_i \leqq y_j$ is now incorporated into the likelihood through the summation over the restricted set $\mathcal{A}$, which recall that was defined as $\mathcal{A} = \{(i,j) \in \{1, \ldots, n\} | y_i \leqq y_j\}$. We use the following prior

$$
\begin{gathered}
p(\theta) = p\left(\sigma_v, \sigma_u, \Sigma, \Omega, \mu\right) \\
\sigma_v^{-(\bar{n}_v+1)} e^{-\bar{q}_v/(2\sigma_v^2)} \cdot \sigma_u^{-(\bar{n}_u+1)} e^{-\bar{q}_u/(2\sigma_u^2)} \cdot \\
|\Sigma|^{-(2K+\bar{n}_\Sigma+1)/2} \exp\left\{ -\frac{1}{2}\bar{A}_\Sigma \Sigma^{-1} \right\} \cdot |\Omega|^{-(2K+\bar{n}_\Omega+1)/2} \exp\left\{ -\frac{1}{2}\bar{A}_\Omega \Omega^{-1} \right\},
\end{gathered} \tag{11}
$$

where $\bar{n}_v, \bar{n}_u, \bar{q}_v, \bar{q}_u, \bar{n}_\Sigma, \bar{n}_\Omega$, are prior non-negative parameters, $\bar{A}_\Sigma$ and $\bar{A}_\Omega$ are $(2K) \times (2K)$ positive semi-definite matrices containing prior parameters.[5] Our prior on μ is $\mathcal{N}(0, h_\mu^2 \mathrm{I})$ with h_μ =100, which is proper but diffuse. Our benchmark prior is proper but diffuse through the specification of the following parameters:

$$
\bar{n}_v = \bar{n}_u = \bar{n}_\Sigma = \bar{n}_\Omega = \bar{n} = 1, \tag{12}
$$

[5]These priors are standard in Bayesian analysis (e.g., Zellner, 1971, p. 371, equations (A.37a) and (A.37b); p. 395, equation (B.71), and for actual use, see p. 225, p. 242).

$$\bar{q}_v = \bar{q}_u = \bar{q} = 0.001, \tag{13}$$

$$\bar{A}_\Sigma = \bar{A}_\Omega = \bar{q} I_{2K}, \tag{14}$$

where I_d is the identity matrix in $\mathbb{R}^{d \times d}$.

To provide access to the posterior in (10), we can use MCMC. The basic building blocks are drawing random numbers from the full posterior conditional distributions (see Technical Appendix A).[6]

The most important question is how to evaluate the hypothesis that a given data set is c-rationalizable. Varian's (1985) approach was essentially to use the "test statistic" $\frac{1}{\sigma_e^2} \sum_{i=1}^{n} \sum_{k=1}^{K} (x_{ik} - \chi_{ik})^2$ in relation to (2). As he noted

> "We are simply asking for the minimal perturbation of the data that satisfies WACM. If the minimal perturbation is small relative to the amount of noise thought to be present in the data then it seems reasonable to accept the null hypothesis." (op. cit., p. 449).

Part of our approach to WACM is that in relation to (8) we should have $\lambda = \frac{\sigma_u}{\sigma_v}$ "large enough", and $\Lambda^{-1} = \frac{|\Sigma|}{|\Omega|}$ to be "small enough". The former is the ratio of "signal" (measured by the variance of inefficiency) to noise measured by its variance (see (8)) and the latter is a ratio of generalized variances (the determinants of covariance matrices). If Λ^{-1} is "small enough" then the deviations of x_i from χ_i (and w_i from ω_i) should be small relative to the variations in the latent variables, captured by Ω. If we define for simplicity $\Lambda = \frac{|\Omega|}{|\Sigma|}$ then both Λ and λ should be "large enough".

[6] In all cases we use 150,000 MCMC iterations omitting the first 50,000 to mitigate the impact of possible start up effects.

2.2 Pair-specific variances

If we let $\sigma^2_{v,ij}$ and $\sigma^2_{u,ij}$ represent pair specific scale parameters in (8), then the posterior in (10) becomes:

$$\begin{gathered} p(\theta, \chi, \omega, u; \mathcal{D}) \propto \\ |\Sigma|^{-n/2} \exp \left\{ -\tfrac{1}{2} \left[\sum_{i=1}^{n} (x_i - \chi_i)' \quad (w_i - \omega_i)' \right] \Sigma^{-1} \left[\begin{array}{c} x_i - \chi_i \\ \\ w_i - \omega_i \end{array} \right] \right\} \cdot \\ \sigma_v^{-N} \exp \left\{ -\tfrac{1}{2} \sum_{(i,j)\in\mathcal{A}} \tfrac{1}{\sigma^2_{v,ij}} \left(\omega_i' \chi_i - \omega_i' \chi_j + u_{ij} \right) \right\} \cdot \\ \sigma_u^{-N} \exp \left\{ -\tfrac{1}{2} \sum_{(i,j)\in\mathcal{A}} \tfrac{1}{\sigma^2_{v,ij}} u^2_{ij} \right\} \cdot \\ |\Omega|^{-n/2} \exp \left\{ -\tfrac{1}{2} \sum_{i=1}^{n} \left[(x_i - \mu_\chi)' \ (w_i - \mu_\omega)' \right] \Omega^{-1} \left[\begin{array}{c} x_i - \mu_\chi \\ \\ w_i - \mu_\omega \end{array} \right] \right\} \cdot p(\theta). \end{gathered} \tag{15}$$

The different scale parameters are a priori distributed as

$$\sigma^2_{v,ij} \sim \sigma^2_v, \qquad\qquad \sigma^2_{u,ij} \sim \sigma^2_u, \tag{16}$$

where σ^2_v and σ^2_v have the same interpretation and priors as before. In this particular case we may want to adopt proper priors changing slightly (12) to

$$\bar{n}_v, \bar{n}_u > 1 \tag{17}$$

for example, $\bar{n}_v = \bar{n}_u = 1.1$.

2.3 Output measured with error

The assumption that output is measured without error is not reasonable when inputs and prices are subject to such errors. Let us modify (5) to the following.

$$\begin{bmatrix} x_i \\ w_i \\ y_i \end{bmatrix} = \begin{bmatrix} \chi_i \\ \omega_i \\ \psi_i \end{bmatrix} + \epsilon_i, \ \epsilon_i \sim \text{i.i.d. } \mathcal{N}_{2K+M}\left(\mathbf{0}, \Sigma)\right), \tag{18}$$

where we keep the symbol for Σ in the interest of simplicity, subject to the optimality restrictions

$$\omega_i'\chi_i = \omega_i'\chi_j + v_{ij} - u_{ij}, \ \forall (i,j) \in \mathscr{A}, \tag{19}$$

where $\mathscr{A} = \{(i,j) : \psi_i \leqq \psi_j\}$, and v_{ij}, u_{ij} are as in the previous subsection. The problem is that now the set $\mathscr{A}$ cannot be defined in advance as ψ_i represents latent unobserved output. We also define $\mathscr{A}_i = \{j : \psi_i \leqq \psi_j\}$. Bayesian numerical inference by MCMC is not entirely different in this case except, of course, for the fact that we have n additional latent variables $\{\psi_i\}_{i=1}^n$. The number of different pairs is denoted, again, by N for simplicity. There are, certainly, "shortcuts" that bypass the problem of the additional latent variable and its involvement in the optimality conditions. One such idea is to treat ψ_is independently of χ_i and ω_i, assuming, for example, that $y_i \sim$ i.i.d. $\mathcal{N}(\mu_y, \sigma_y^2)$ and set

$$\mathcal{A} = \{(i,j) : \Pr(y_i \leq y_j) \geq 1 - a\}, \tag{20}$$

for a certain small number a, for example $a = 0.05$. Effectively this converts the deterministic inequalities $y_i \leqq y_j$ to probabilistic inequalities which, however, can be computed in advance; even where output follows a different distribution (log-normal, mixture of log-normal distributions, etc.). In this case we do not have the latent output variables $\{\psi_i\}_{i=1}^n$ at all.

The approach in (18) is, perhaps, more reasonable as it allows for correlated output, prices and input quantities. Details of Bayesian numerical inference by MCMC are available as a separate Appendix.

3 Empirical Application

We have an unbalanced panel with 2,397 bank-year observations for 285 large U.S. commercial banks operating in 2001-2010, whose total assets were more than one billion dollars (in 2005 U.S. dollars) in the first three years of observation. We use the data for 2009 to focus on the effects of the financial crisis. The data come from Call Reports available from the Federal Reserve Bank of Chicago. This data set has been used in Malikov et al. (2016), where more detailed description of the data construction can be found (see their Section 5).

The list of included variables is as follows: y_1 - Consumer Loans, y_2 - Real Estate Loans, y_3 - Commercial & Industrial Loans, y_4 - Securities, y_5 - Off-Balance Sheet Activities Income, x_1 - Labor, number of full-time employees, x_2 - Physical Capital (Fixed Assets), x_3 - Purchased Funds, x_4 - Interest-Bearing Transaction Accounts, x_5 - Non-Transaction Accounts. For simplicity, we aggregate all outputs to one using their revenue shares as weights.[7]

[7]Extending the WACM to the case of multiple outputs is, of course, possible but the requirement is $y_i \leqq y_j$ in the sense that firm i produces weakly less than firm j for all outputs would leave too few observations to compare to, if any at all.

3.1 Basic Model

Figure 1: Bivariate posterior densities of λ and Λ

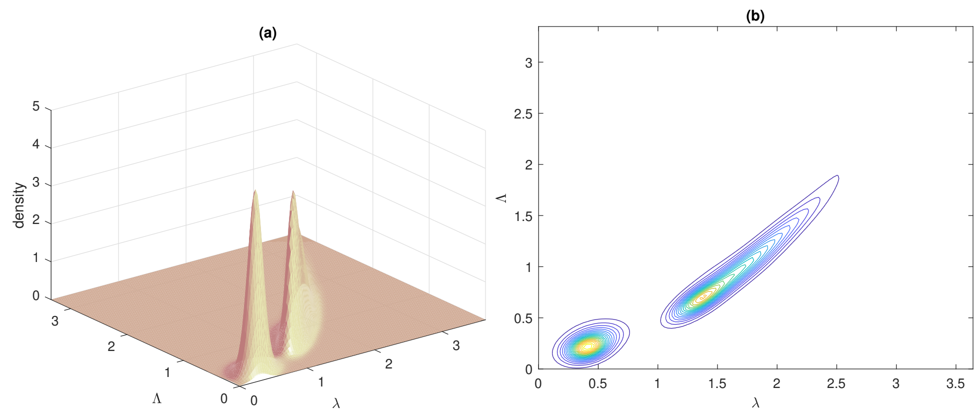
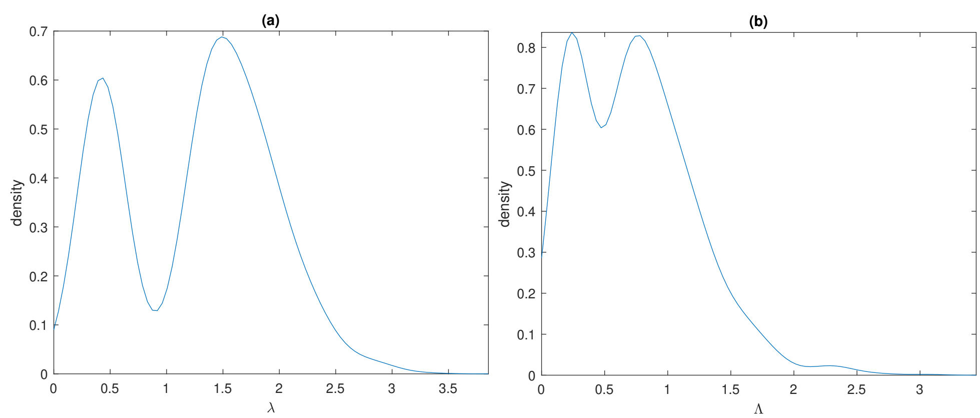


Figure 2: Marginal posterior densities of λ and Λ



From the evidence in Figures 1 and 2, where the aspects of the sample distributions of posterior means of λ and Λ are reported, it turns out that there are two groups in the data, evidenced by the bimodality of the distributions. The first group has values of λ near 0.4 (ranging, approximately, from 0.2 to 0.7) and associated values of Λ close to 0.25. The second group has a positive relationship between λ and Λ which average close to 1.5 and 0.9, respectively, although values such as 2.5 and 2 are plausible as well.

A formal test of the hypothesis $H : \lambda = \Lambda = 0$ can be developed by using the Bayes factor of H given the data. The Bayes factor can be computed by the Savage-Dickey ratio (Dickey, 1971; see also Verdinelli and Wasserman, 1995). The Bayes factor in favor of H can be computed as

$$B = \frac{\int p(\lambda=0,\Lambda=0,\vartheta|\mathcal{D})\,\mathrm{d}\vartheta}{\int p(\lambda=0,\Lambda=0,\vartheta)\,\mathrm{d}\vartheta}, \tag{21}$$

where B is the Bayes factor in favor of H, ϑ denotes parameters and latent variables other than λ and Λ and the denominator is the marginal prior. The numerator may be accurately approximated by

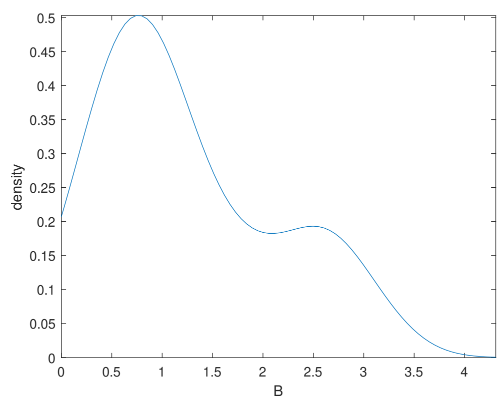
$$\begin{gathered}\int p(\lambda = 0, \Lambda = 0, \vartheta|\mathcal{D})\,\mathrm{d}\vartheta = \int p(\lambda = 0, \Lambda = 0|\vartheta, \mathcal{D})p(\vartheta|\mathcal{D})\,\mathrm{d}\vartheta \simeq \\ S^{-1}\sum_{s=1}^{S} p(\lambda = 0, \Lambda = 0|\vartheta^{(s)}, \mathcal{D}),\end{gathered} \tag{22}$$

where $\left\{\vartheta^{(s)}, s = 1, \ldots, S\right\}$ is a MCMC sample for ϑ. This multivariate integral can be evaluated pointwise for all $\vartheta^{(s)}$. The normalizing constant can be computed using the Laplace approximation (Raftery, 1995).

The integral in the denominator can also be computed relatively easily as $\int p(\lambda = 0, \Lambda = 0, \vartheta)\,\mathrm{d}\vartheta = \int p(\lambda = 0, \Lambda = 0|\vartheta)p(\vartheta)\,\mathrm{d}\vartheta = p(\lambda = \Lambda = 0)$. To compute the prior probability density evaluated at $\lambda = \Lambda = 0$ we draw a sample of size 100,000 from the prior of $\sigma_v, \sigma_u, \Sigma, \Omega$ and compute the corresponding draws of λ and Λ. In turn, the prior probability is computed using a kernel density estimator of their bivariate density.

Under our benchmark prior, the Bayes factor turns out to be close to zero. However, the question is whether there are other priors that can c-rationalize the data. From (12)-(14) we select different values for $\bar{n}$ and $\bar{q}$ and examine the distribution of the Bayes factor, B, across the different values for $\bar{n}$ and $\bar{q}$. Specifically, we draw 10,000 different values for $\bar{n}$ and $\bar{q}$ in the range $[0, 100]$ for both parameters, we repeat MCMC using parallel computation and we compute the approximation of B in (21). The resulting distribution is presented in Figure 3.

Figure 3: Distribution of Bayes factors for c-rationalization across different priors



From the evidence in Figure 3, c-rationalizability in H does not appear plausible in the light of the data as the Bayes factors do not exceed 5 (values greater than ten would be required, see Kass and Raftery, 1995). Our techniques can be extended to profit maximization rationalizability (p-rationalizability) and, in fact, we do not have complications arising from sets $\mathcal{A}$ and $\mathcal{A}_i$ as in (10). Surely, the evidence against c-rationalization does not leave much hope for p-rationalization but we proceed in the interest of illustrating the new techniques and also because c-rationalizability was tested under the assumption that we have a single output which may be restrictive.

From the WAPM in (4) we have the data generating process

$$\begin{bmatrix} \mathbf{y}_i \\ \mathbf{p}_i \end{bmatrix} = \begin{bmatrix} \boldsymbol{\psi}_i \\ \boldsymbol{\pi}_i \end{bmatrix} + \mathbf{V}_i,\ \mathbf{V}_i \sim \mathcal{N}_{2(K+M)}(\mathbf{0}, \boldsymbol{\Sigma}),\ i = 1, \ldots, n, \tag{23}$$

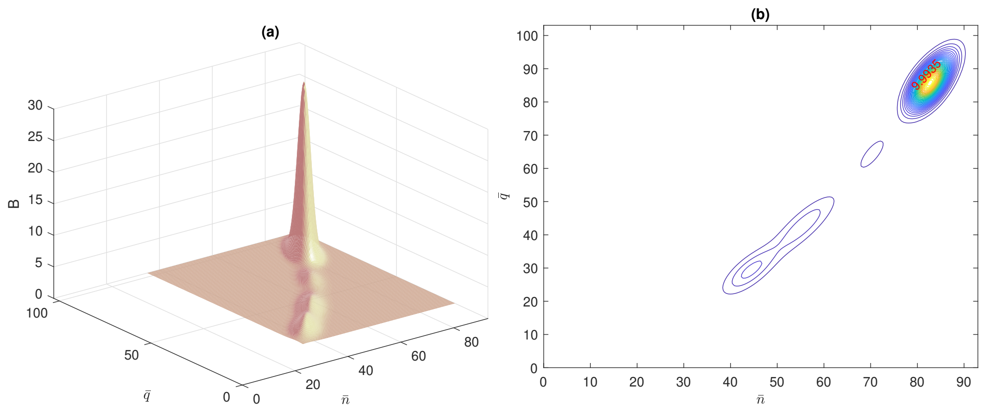
$$\begin{bmatrix} \boldsymbol{\psi}_i \\ \boldsymbol{\pi}_i \end{bmatrix} \sim \mathcal{N}_{2(K+M)}(\boldsymbol{\mu}, \boldsymbol{\Omega}),\ i = 1, \ldots, n, \tag{24}$$

where $\boldsymbol{\mu} = \begin{bmatrix} \mu_\psi \\ \mu_\pi \end{bmatrix}$, along with the "prior restrictions"

$$\boldsymbol{\pi}_i' \boldsymbol{\psi}_i = \boldsymbol{\pi}_i' \boldsymbol{\psi}_j + v_{ij} + u_{ij},\ i \neq j = 1, \ldots, n, \tag{25}$$

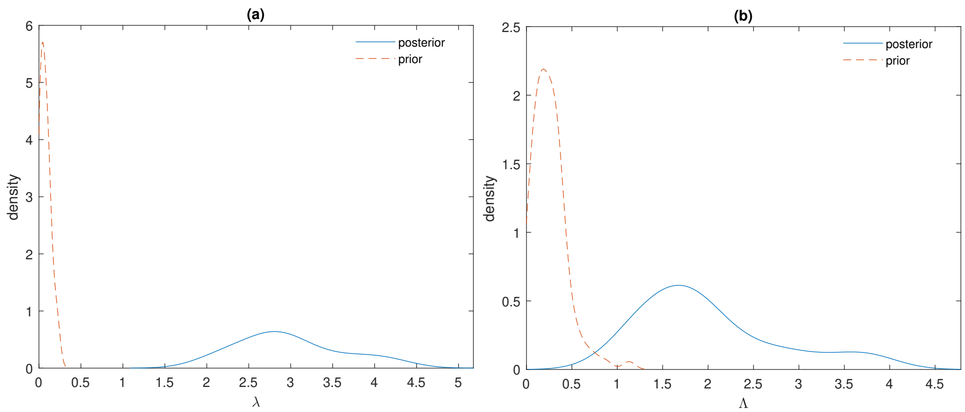
where v_{ij} and $u_{ij} \in \mathbb{R}_+$ have the same interpretation as before and $\boldsymbol{\psi}_j$ is technically feasible when $\boldsymbol{\psi}_i$ was chosen. Again, we let N denote the number of different pairs. Despite the fact that we have $N = n(n-1)$ such restrictions, use of a stochastic frontier model is still possible in moderate to large sample sizes. The distributions of the Bayes factor in favor of p-rationalizability are reported in Figure 4.

Figure 4: Distribution of the Bayes factors for p-rationalization across different priors



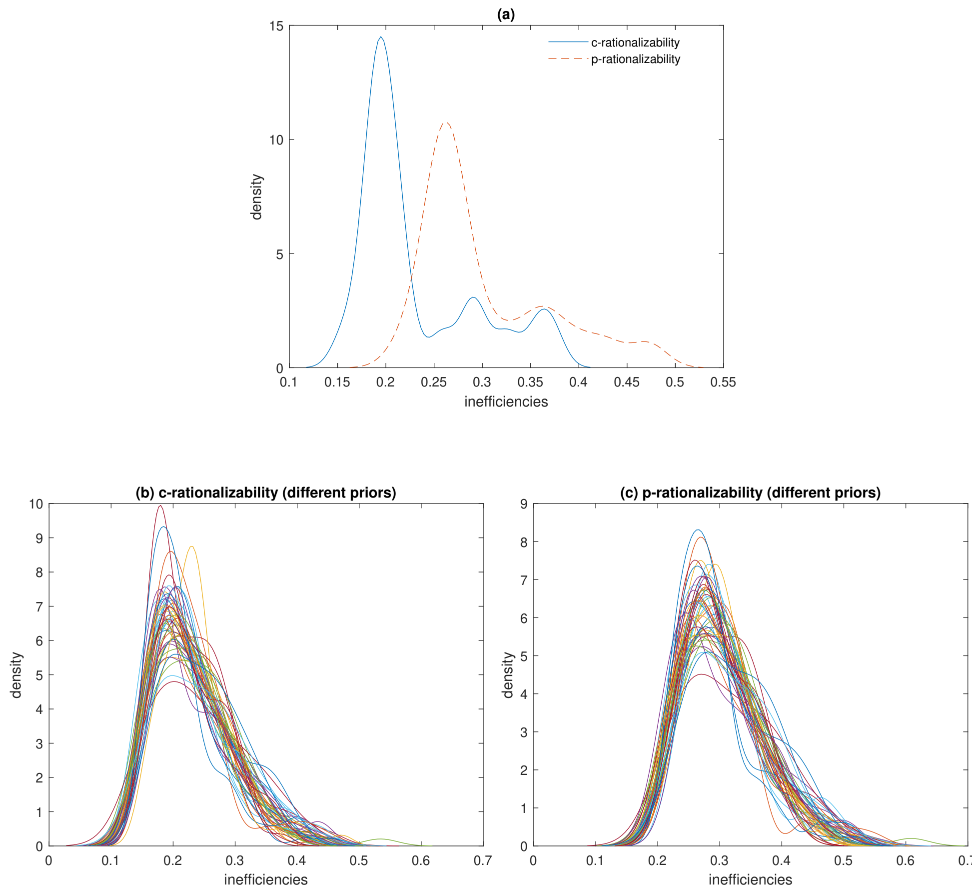
For most banks, the Bayes factors in favor of p-rationalizability are clustered towards small values. For more informative priors (larger values of $\bar{n}$ and $\bar{q}$) the evidence shows that the Bayes factors in favor of p-rationalizability are close to 10 and can be as large as 27. Although the results are sensitive to the prior, larger values of $\bar{n}$ and $\bar{q}$ are relatively plausible as they are diffuse relative to the likelihood. Evidence that corroborates this claim is provided in Figure 5 where we report marginal posterior densities of λ and Λ for p-rationalizability along with their (implied) priors.

Figure 5: Marginal posterior densities of λ and Λ for p-rationalizability



Of course, it is of central interest to measure inefficiencies or departures from c- or p-rationalization such as the u's in (8). Sample distributions of the posterior mean estimates of departures from c- and p-rationalizability are reported in panel (a) of Figure 6. Under fifty randomly chosen alternative priors (see Technical Appendix B) the results are presented in panels (b) and (c) and seem fairly robust to the priors. On average, the departures from c-rationalizability are close to 25% and range from 10% to 40%, approximately, whereas the departures from p-rationalizability are close to 33% and range from 15% to slightly less than 55%, approximately. With other priors, much higher upper bounds are possible, for example 60% and nearly 70% for c- and p-rationalizability, respectively.

Figure 6: Estimated departures from c- and p- rationalizability



3.2 Log-normality

The assumption of normality in (6) may not be as reasonable as the assumption of log–normality particularly when there is skewness and other departures from normality. In these cases, we can replace (6) or (24) by

$$\begin{bmatrix} \ln \boldsymbol{\psi}_i \\ \ln \boldsymbol{\pi}_i \end{bmatrix} \sim \mathcal{N}_{2(K+M)}(\boldsymbol{\mu}, \boldsymbol{\Omega}),\ i = 1, \ldots, n, \tag{26}$$

along with (23). In this case, we have to replace the posterior by the following:

$$
\begin{aligned}
& p(\vartheta, \boldsymbol{\psi}, \boldsymbol{\pi}, \mathbf{u}|\mathcal{D}) \propto \\
& |\boldsymbol{\Sigma}|^{-n/2} \exp\left\{-\tfrac{1}{2}\textstyle\sum_{i=1}^{n}(\mathbf{a}_i - \boldsymbol{\alpha}_i)'\boldsymbol{\Sigma}^{-1}(\mathbf{a}_i - \boldsymbol{\alpha}_i)\right\} \cdot \\
& |\boldsymbol{\Omega}|^{-n/2} \exp\left\{-\tfrac{1}{2}\textstyle\sum_{i=1}^{n}(\ln \boldsymbol{\alpha}_i - \boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\ln \boldsymbol{\alpha}_i - \boldsymbol{\mu}) - \sum_{i=1}^{n} \ln \boldsymbol{\alpha}_i\right\} \cdot \\
& \sigma_v^{-N} \exp\left\{-\tfrac{1}{2\sigma_v^2}\textstyle\sum_{i\neq j}\left(\boldsymbol{\pi}_i'\boldsymbol{\psi}_i - \boldsymbol{\pi}_i'\boldsymbol{\psi}_j - u_{ij}\right)^2\right\} \cdot \sigma_u^{-N} \exp\left\{-\tfrac{1}{2\sigma_u^2}\textstyle\sum_{i\neq j} u_{ij}^2\right\} \cdot p(\vartheta),
\end{aligned}
\tag{27}
$$

where $\mathbf{a}_i = [\mathbf{y}_i', \mathbf{p}_i']'$, $\boldsymbol{\alpha}_i = [\boldsymbol{\psi}_i', \boldsymbol{\pi}_i']'$, and $\mathbf{u} = [u_{ij}, i \neq j = 1, \ldots, n]$. The presence of logs creates technical problems for implementing the Gibbs sampler but we can approximate $\ln \boldsymbol{\alpha}_i \simeq \ln \boldsymbol{\alpha}_i^{(s-1)} + \boldsymbol{\alpha}_i/\boldsymbol{\alpha}_i^{(s-1)} - 1$, where $\boldsymbol{\alpha}_i^{(s-1)}$ is the value of $\boldsymbol{\alpha}_i$ in the previous MCMC iteration and the division is understood in the element-wise sense. With this modification we can use the techniques in Technical Appendix A. These MCMC draws are accepted, eventually, with a Metropolis-Hastings probability involving the ratio of an exact posterior in (27) and an approximate posterior as in Kim, Shephard, and Chib (1998). Across priors and different initial conditions the acceptance rate of this procedure was well over 90%. Results from the lognormal data generating process are reported in Figure 7.

Figure 7: Aspects of p-rationalization under lognormality

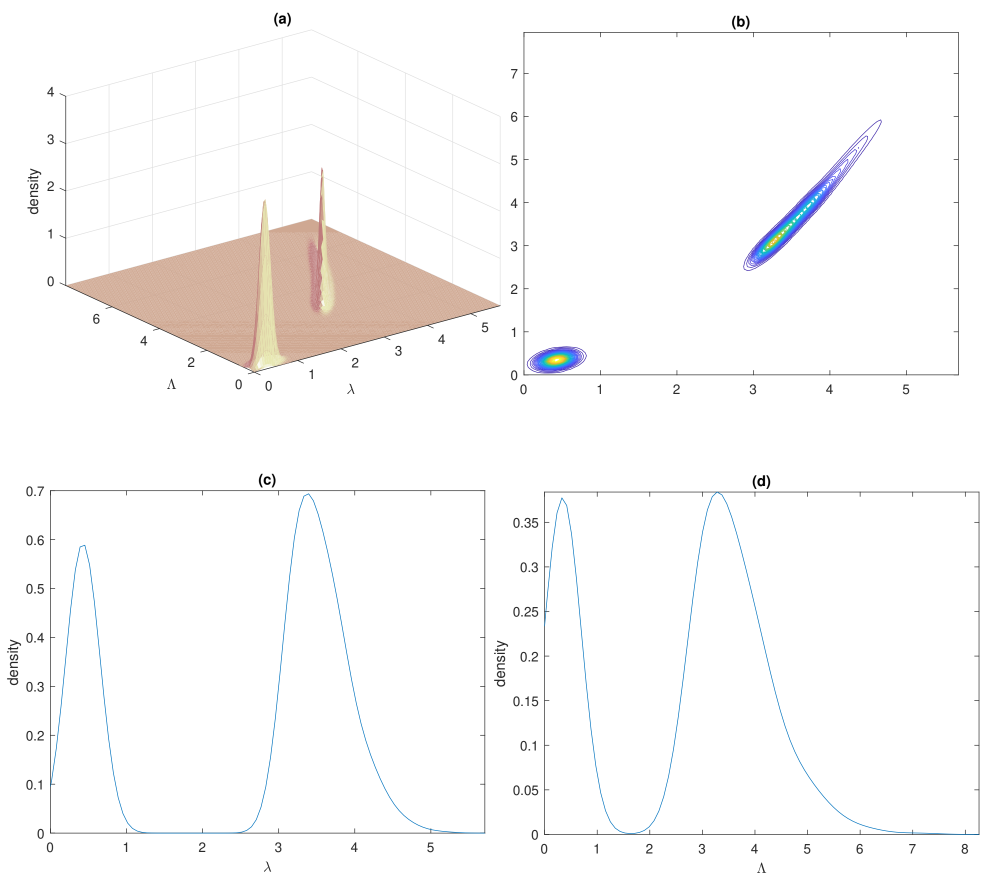
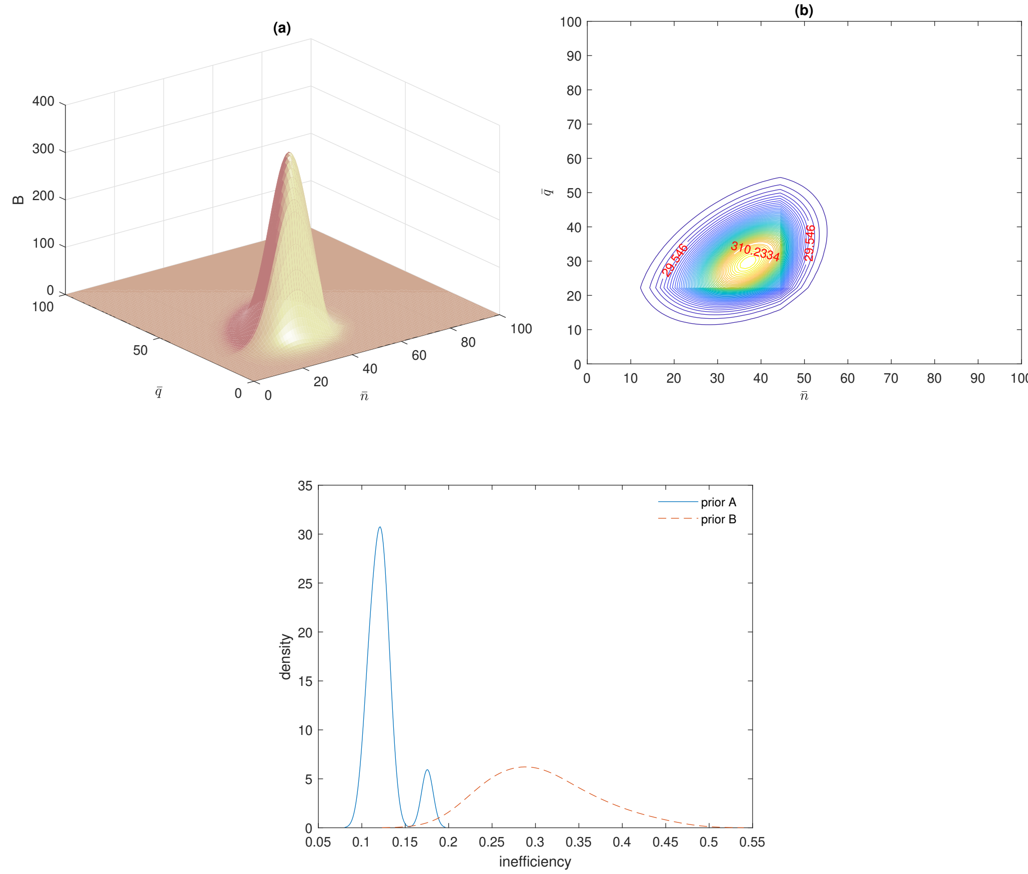


Figure 8: Bayes factors for p-rationalizability under lognormality



The evidence in Figures 7 and 8 shows that there is a range of prior parameters roughly around $\bar{n} = \bar{q} = 30$ which provide large Bayes factors in favor of p-rationalization. For very low or very large values of these parameters the Bayes factors turn out to be quite small. So, overall, there is some limited evidence in favor of p-rationalizability although this evidence depends on prior beliefs. In panel (c) of Figure 8, we report the sample distribution of posterior mean departures (inefficiency) from p-rationalizability. Prior A has $\bar{n} = \bar{q} = 30$ and prior B has $\bar{n} = \bar{q} = 1$ to discriminate between the cases shown in panels (a) and (b). Clearly, departures from p-rationalizability are much smaller around the values of the prior parameters that provide evidence in favor of p-rationalizability, providing corroborating evidence that these priors are compatible with p-rationalization of the data.

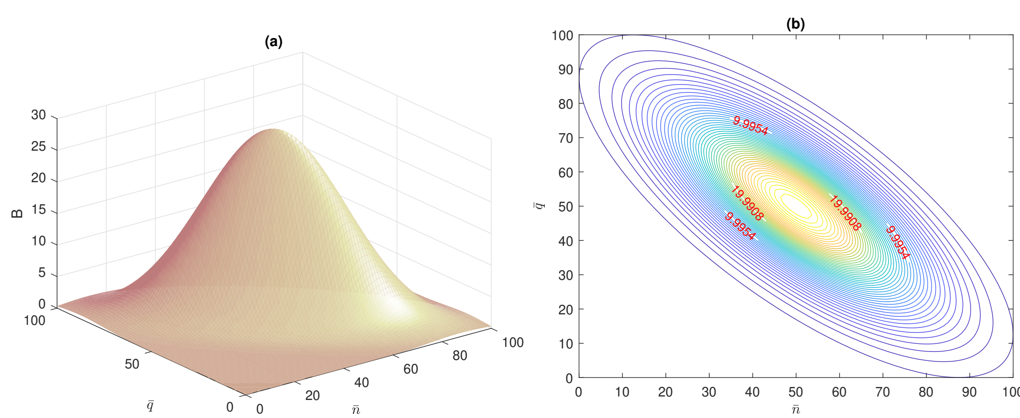
3.3 Removing the half-normality assumption on p-optimality deviations

The assumption that u_{ij}s follow a half-normal distribution may seem as rather restrictive. Instead, we can adopt a modified formulation in the spirit of Cornwell, Schmidt and Sickles (1990, CSS). Specifically, we set

$$u_{it} = \exp\left\{\sum_{j=0}^{P} a_{ij} t^j\right\}, \tag{28}$$

where $a_{i0}, a_{i1}, \ldots, a_{iP}$ are bank-specific coefficients to be estimated. Therefore, departures from p-rationalization are modeled as flexible polynomials of time t. We introduce here the exponential transformation so that all deviations are non-negative. Based on a marginal likelihood we select $P = 2$. To avoid the proliferation of bank-specific parameters we adopt a standard normal prior. The new evidence is provided in Figure 9.

Figure 9: Bayes factors for p-rationalization (CSS-like formulation)



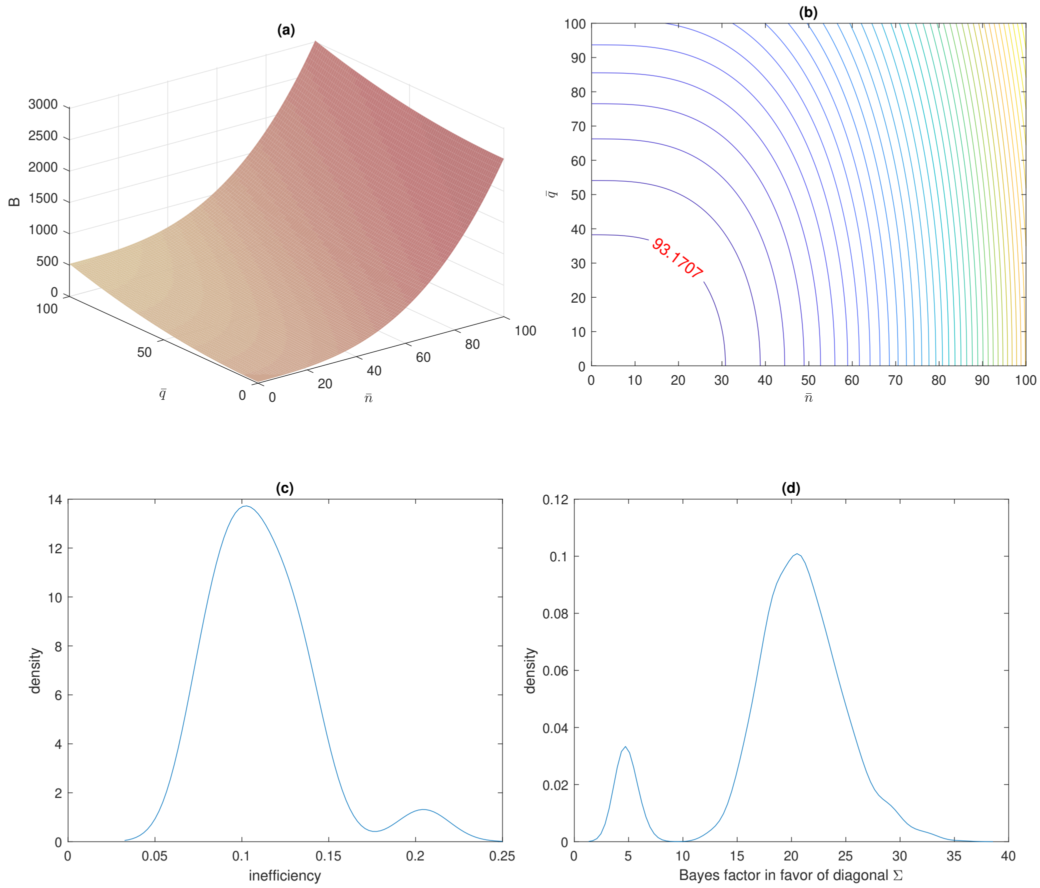
Under the CSS-like formulation, the evidence in favor of p-rationality is considerable for larger values of $\bar{n}$ and $\bar{q}$ so this evidence is not incompatible with previous findings, in a qualitative sense. Bayes factors range from 10 to 25 for these priors but, again, the evidence is not unanimous across priors.

The marginal evidence in favor of p-rationality owes much to the fact that we allow $\mathbf{\Sigma}$ to be a general matrix. This matrix relates to measurement error. One possibility is to proceed along the lines of Varian (1985) and assume that $\mathbf{\Sigma}$ is a diagonal matrix with the diagonal elements h_{ii} having the prior

$$p(h_{ii}) \propto h_{ii}^{-(\bar{n}+1)} e^{-\bar{q}/(2h_{ii}^2)}.$$

We proceed using the CSS-like formulation and we provide the new Bayes factors in Figure 10.

Figure 10: Bayes factors for p-rationalization (CSS-like formulation and diagonal $\mathbf{\Sigma}$)



The restriction that measurement errors have a diagonal covariance matrix definitely provides more evidence in favor of p-rationalization (panels (a) and (b)). Deviations from p-rationalization are provided in panel (c) for $\bar{n} = \bar{q} = 10$. To conclude this sub-section, we note that a lot depends on the data generating process as well as the assumptions that we make about measurement errors. Allowing for a general matrix $\mathbf{\Sigma}$ considerably weakens the evidence in favor of p-rationality, while assuming a diagonal covariance of measurement errors, increases considerably the Bayes factor. To examine whether a common measurement error variance is reasonable in the light of the data, we compute the Bayes factor in favor of a diagonal $\mathbf{\Sigma}$ matrix, against the hypothesis of a general $\mathbf{\Sigma}$ matrix for various values of $\bar{n}$ and $\bar{q}$. Using 1,000 different draws, the distribution of Bayes factors is presented in panel (d) of Figure 10. From the evidence in panel (d) it appears that, for the most part, a diagonal $\mathbf{\Sigma}$ is supported in the light of the data so, profit maximization is not highly unlikely.

3.4 A mixture-of-lognormal-distributions formulation

The specification in (27) may be somewhat restrictive in that there may be several groups in the data. In the interest of generalizing the model we adopt a multivariate mixture-of-lognormal-distributions:

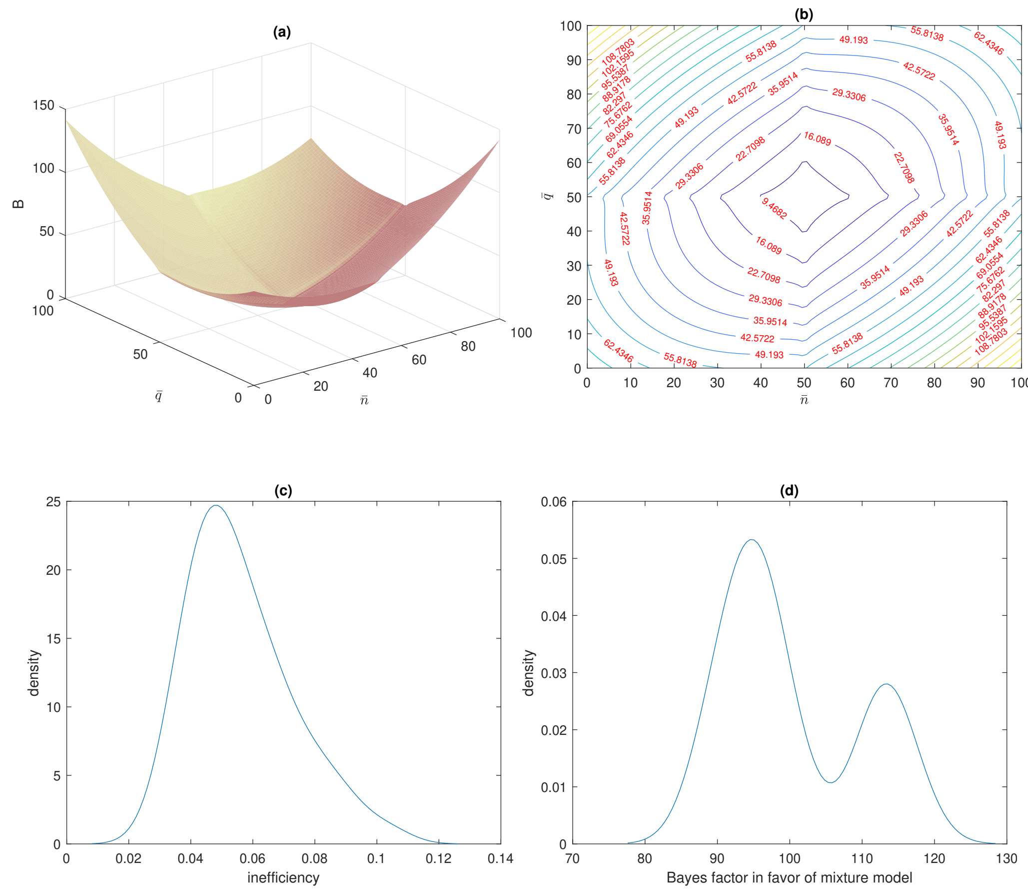
$$\begin{bmatrix} \ln \boldsymbol{\psi}_i \\ \ln \boldsymbol{\pi}_i \end{bmatrix} \sim \mathcal{N}_{2(K+M)}(\boldsymbol{\mu}^{(g)}, \mathbf{\Omega}^{(g)}), \text{ with probability } \Pi_g,\ i = 1, \ldots, n, \tag{29}$$

where $g \in \{1, \ldots, G\}$ denotes group membership with an upper bound G, and Π_g denotes probability of group g. Evidence based on marginal likelihoods shows that the Bayes factor in favor of $G = 3$ is 255.32 (22.30 for $G = 2$ and 1.18 for $G = 4$). Our priors for $\left\{\boldsymbol{\mu}^{(g)}, \mathbf{\Omega}^{(g)}, g = 1, \ldots, G\right\}$ are the same as the priors of $\boldsymbol{\mu}$ and $\mathbf{\Omega}$ in the previous section, and we adopt a flat prior for probabilities $\Pi^{(g)}$s in the unit simplex.[8] Our MCMC methods can be easily extended to the case of a multivariate mixture-of-normal-distributions.

Bayes factors as a function of the prior parameters $\bar{n}$ and $\bar{q}$ in Figure 11 provide more support for p-optimality even at relatively low values of these parameters which provide diffuse priors. This finding is, apparently, due to allowing for a multivariate-mixture-of-lognormal-distributions which allowed for more heterogeneity relative to (26). A diagonal $\mathbf{\Sigma}$ is still supported by the data although we have been unable to find evidence to support diagonal $\mathbf{\Omega}^{(g)}$ matrices. Surprisingly, deviations from p-rationalizability constraints are relatively small (see their sample density in panel (c) of Figure (11)

[8] Detailed results are available on request.

Figure 11: Bayes factors for p-rationalization (multivariate-mixture-of-normal distributions)



averaging 5.6%, with a median of 5.3%, and standard deviation 1.64%). Moreover, the multivariate-mixture-of-normal-distributions has an overwhelming Bayes factor against the specification in the previous section. Across the different priors, the distribution is reported in panel (d) of Figure 11.

4 Concluding remarks

In this paper we propose operational versions of tests for the weak axioms of cost minimization and profit maximization to c- or p-rationalize observed data sets measured with noise in both prices and quantities. For many years since Varian (1985) this has been a difficult task without a satisfactory statistical solution. We operationalize the c- or p-rationalizability constraints in the form of prior constraints among the otherwise unobservable latent price and quantity data, and we employ formal Bayesian methods organized around MCMC to i) test c- or p-rationalizability, ii) measure the extent of deviations from c- or p-rationalizability. Formal testing is based on obtaining Bayes factors in favor of the hypothesis that the data can be c- or p-rationalized.

We have proposed various ways to examine the robustness of the basic specification that can be quite useful in practice, including a normal and lognormal specification of the data generating process, a flexible model for firm-specific deviations from p-optimality as well as a multivariate-mixture-of-(log)normal-distributions for the data generating process of the unobservable, latent, price and quantity data.

Our results indicate that flexible data generating processes need to be specified along with simplifications (like, for example, diagonal covariance matrices) which can be substantiated using Bayes factors for a variety of priors. A natural avenue for future research is an adaptation of the proposed approach to the consumer context.

References

[1] Afriat, S.N. (1967). The Construction of a Utility Function From Expenditure Data. International Economic Review 8, 67–77.

[2] Afriat, S. N., (1972). Efficiency Estimation of Production Functions. International Economic Review 13:3, pp. 568-598.

[3] Aigner, D., Lovell, C.A.K., and Schmidt, P. (1977). Formulation and Estimation of Stochastic Frontier Production Function Models. Journal of Econometrics 6, 1 (1977), pp. 21--37.

[4] Chavas, J.-P. and Cox, T. L. (1990). A Non-Parametric Analysis of Productivity: The Case of U.S. and Japanese Manufacturing. The American Economic Review 80:3 (1990), pp. 450--464.

[5] Cornwell, C., P. Schmidt, and R. Sickles (1990). Production Frontiers with Cross-Sectional and Time-series Variation in Efficiency Levels. Journal of Econometrics 46, 185–200.

[6] Dickey, J. (1971). The Weighted Likelihood Ratio, Linear Hypotheses on Normal Location Parameters. The Annals of Statistics 42, 204–223.

[7] Du, P., Parmeter, C. F., and Racine, J. S. (2013). Nonparametric Kernel Regression With Multiple Predictors and Multiple Shape Constraints. Statistica Sinica, 23, 1347–1371.

[8] Echenique, F., Lee, S., Shum, S. (2011). The Money Pump as a Measure of Revealed Preference Violations. Journal of Political Economy 119 (6), 1201–1223.

[9] Färe, R. and Grosskopf, S., (1995). Nonparametric Tests of Regularity, Farrell efficiency, and Goodness-of-fit. Journal of Econometrics 69:2, pp. 415-425.

[10] Fleissig, A.R., Whitney, G.A. (2003). A New PC-Based Test for Varian's Weak Separability Conditions. Journal of Business & Economic Statistics 21 (1), 133–144.

[11] Hanoch, G. and Rothschild, M., (1972). Testing the Assumptions of Production Theory: A Non-parametric Approach. The Journal of Political Economy 80:2, pp. 256--275.

[12] Geweke, J. (1992). Evaluating the Accuracy of Sampling-based Approaches to Calculating Posterior Moments. In *Bayesian Statistics 4* (J. M. Bernado, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Clarendon Press, Oxford, UK, 169–193.

[13] Geweke, J. and Keane, M. (2007). Smoothly Mixing Regressions. Journal of Econometrics 138, 252–290.

[14] Keshvari, A., & Kuosmanen, T. (2013). Stochastic Non-convex Envelopment of Data: Applying Isotonic Regression to Frontier Estimation. European Journal of Operational Research, 231(2), 481–491.

[15] Kass, R.E., & Raftery, A.E. (1995). Bayes Factors. Journal of the American Statistical Association 90 (430), 773–795.

[16] Kim, S., N. Shephard, and S. Chib (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. Review of Economic Studies 65, 361–393.

[17] Kuosmanen, T. (2008). Representation Theorem for Convex Nonparametric Least Squares. Econometrics Journal, 11(2), 308–325.

[18] Kuosmanen, T. & Kortelainen, M. (2012). Stochastic Non-smooth Envelopment of Data: Semi-parametric Frontier Estimation Subject to Shape Constraints. Journal of Productivity Analysis, 38 (1), 11–28.

[19] Malikov, E., S. C. Kumbhakar, & M. G. Tsionas, 2016. A Cost System Approach to the Stochastic Directional Technology Distance Function with Undesirable Outputs: The Case of U.S. Banks in 2001-2010. Journal of Applied Econometrics, 31 (7), 1407–1429.

[20] Norets, A. (2012). Approximation of Conditional Densities by Smooth Mixtures of Regressions, Annals of Statistics 38(3), 1733–1766.

[21] Parmeter, C. F. and Zelenyuk, V. (2019). Combining the Virtues of Stochastic Frontier and Data Envelopment Analysis, Operations Research 67 (6), 1628—1658.

[22] Raftery, A. (1995). Hypothesis Testing and Model Selection Via Posterior Simulation, in Practical Markov Chain Monte Carlo, eds. W. R. Gilks, D. J. Spieglehalter, and S. Richardson, London: Chapman and Hall, London.

[23] Smeulders, B., Crama, Y., Spieksma, F. C. R. (2019). Revealed Preference Theory: An Algorithmic Outlook. European Journal of Operational Research 272 (3), 1 803–815.

[24] Varian, H. (1982a). Nonparametric Methods in Demand Analysis, Economics Letters 9, 23–29.

[25] Varian, H. (1982b). The Nonparametric Approach to Demand Analysis, Econometrica 50, 945-973.

[26] Varian. H., (1982c), Trois evaluations de l'impact 'social' d'un changement de prix. Cahiers du Seminar d'Econometrie 24, 13–30.

[27] Varian, H. (1983a). Nonparametric Tests of Consumer Behavior, Review of Economic Studies 50, 99–110.

[28] Varian, H. (1983b). Nonparametric Tests of Models of Investor Behavior. Journal of Financial and Quantitative Analysis 18, 269–278.

[29] Varian, H. (1984). The Nonparametric Approach to Production Analysis, Econometrica 52, 579-597.

[30] Varian, H. (1985). Non-parametric Analysis of Optimizing Behavior with Measurement Error. Journal of Econometrics 30, 445–458.

[31] Verdinelli, I., & Wasserman, L. (1995). Bayes Factors Using a Generalization of the Savage-Dickey Density Ratio. Journal of the American Statistical Association 90 (430), 614–618.

[32] Yagi, D., Chen, Y., Johnson, A. L., and Kuosmanen, T. (2020). Shape-Constrained Kernel-Weighted Least Squares: Estimating Production Functions for Chilean Manufacturing Industries. Journal of Business & Economic Statistics 38 (1), 43–54.

[33] Zellner, A. (1971). An Introduction to Bayesian Inference in Econometrics. Wiley, New York.

Technical Appendix A

$$u_{ij}|u_{-(ij)}, \theta, \chi, \omega, D \sim \mathcal{N}_+\left(\tilde{u}_{ij}, s_u^2\right), \tag{A.1}$$

where $\tilde{u}_{ij} = \frac{-r_{ij}\sigma_u^2}{\sigma_v^2+\sigma_u^2}$, $s_u^2 = \frac{-\sigma_v^2\sigma_u^2}{\sigma_v^2+\sigma_u^2}$, and $r_{ij} = \omega_i'\chi_i - \omega_i'\chi_j$.

$$\tfrac{\bar{q}_u+\sum_{(i,j)\in A} u_{ij}^2}{\sigma_u^2}|\cdot, \mathcal{D} \sim \chi^2_{N+\bar{n}_u}, \tag{A.2}$$

viz. a *chi*-square distribution (the symbol is not to be confused with χ_i).

$$p(\Sigma|\cdot, \mathcal{D}) \propto |\Sigma|^{-(n+2K+\bar{n}_\Sigma+1)/2} \exp\left\{-\tfrac{1}{2}\mathbf{A}_\Sigma\Sigma^{-1}\right\}, \tag{A.3}$$

where $\mathbf{A}_\Sigma = \det\left[\sum_{i=1}^n \begin{bmatrix} x_i - \chi_i \\ w_i - \omega_i \end{bmatrix} \begin{bmatrix} x_i - \chi_i \\ w_i - \omega_i \end{bmatrix}' + \bar{A}_\Sigma\right]$,

$$p(\Omega|\cdot, \mathcal{D}) \propto |\Omega|^{-(n+2K+\bar{n}_\Omega+1)/2} \exp\left\{-\tfrac{1}{2}\mathbf{A}_\Omega\Omega^{-1}\right\}, \tag{A.4}$$

where $\mathbf{A}_\Omega = det\left[\sum_{i=1}^n \begin{bmatrix} x_i - \mu_\chi \\ w_i - \mu_\omega \end{bmatrix} \begin{bmatrix} x_i - \mu_\chi \\ w_i - \mu_\omega \end{bmatrix}' + \bar{A}_\Omega\right]$.

$$\tfrac{\sum_{(i,j)\in\mathcal{A}}\left(\omega_i'\chi_i - \omega_i'\chi_j + u_{ij}\right)}{\sigma_v^2}|\cdot, \mathcal{D} \sim \chi^2_N. \tag{A.5}$$

To draw from the conditional posterior distribution of χ_i we notice that we have the following equations for a given $i \in \{1, \ldots, n\}$:

$$\begin{bmatrix} \chi_i \\ \omega_i \end{bmatrix} \sim \text{i.i.d. } \mathcal{N}_{2K}\left(\mu, \Omega\right), \tag{A.6}$$

$$\begin{bmatrix} x_i \\ w_i \end{bmatrix} = \begin{bmatrix} \chi_i \\ \omega_i \end{bmatrix} + \varepsilon_i,\ \varepsilon_i \sim \mathcal{N}_{2K}(\mathbf{0}, \Sigma), \tag{A.7}$$

$$\omega_i' \chi_j - u_{ij} = \omega_i' \chi_i + v_{ij}\ \forall j \in \mathcal{A}_i. \tag{A.8}$$

Since we condition on ω, we can derive

$$\chi_i | \omega_i \sim \mathcal{N}_K \left(\mu_\chi + \Omega_{12}\Omega_{22}^{-1}\Omega_{12}'(\omega_i - \mu_\omega),\ \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{12}' \right), \tag{A.9}$$

instead of (A.6), and

$$\chi_i | \omega_i \sim \mathcal{N}_K \left(x_i + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'(\omega_i - w_i),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}' \right), \tag{A.10}$$

instead of (A.7). Using Theil's mixed estimator we can derive the posterior conditional distribution of χ_i. Let us define

$$\mu_\omega^* = \mu_\omega + \Omega_{12}\Omega_{22}^{-1}\Omega_{12}'(\omega_i - \mu_\omega), \tag{A.11}$$

$$\Omega_{11,*} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{12}', \tag{A.12}$$

$$x_i^* = x_i + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'(\omega_i - w_i), \tag{A.13}$$

$$\Sigma_{11,*} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}', \tag{A.14}$$

$$a_j = \omega_i' \chi_j - u_{ij} = \omega_i' \chi_i + v_{ij}\ \forall j \in \mathcal{A}_i. \tag{A.15}$$

Suppose $\mathbf{a}_i = [a_1, \ldots, a_{N_i}]'$, $\Xi_i = \iota_{n_i} \otimes \omega_i'$. Then the posterior conditional distribution of χ_i is

$$\chi_i | \cdot, \mathcal{D} \sim \mathcal{N}_K \left(\hat{\chi}_i,\ \mathbf{S}_{\chi_i} \right), \tag{A.16}$$

where

$$\hat{\chi}_i = \left(\Omega_{11,*}^{-1} + \Sigma_{11,*}^{-1} + \tfrac{1}{\sigma_v^2}\Xi_i'\Xi_i\right)^{-1}\left(\Omega_{11,*}^{-1}\mu_\omega^* + \Sigma_{11,*}^{-1}x_i^* + \tfrac{1}{\sigma_v^2}\Xi_i'\mathbf{a}_i\right), \tag{A.17}$$

$$\mathbf{S}_{\chi_i} = \left(\Omega_{11,*}^{-1} + \Sigma_{11,*}^{-1} + \tfrac{1}{\sigma_v^2}\Xi_i'\Xi_i\right)^{-1}. \tag{A.18}$$

The result for the the posterior conditional distribution of ω_i is analogous and so we omit it here for the sake of brevity.

Technical Appendix B

In this Appendix we provide evidence of MCMC performance. We are especially interested in autocorrelation functions (acf) and relative numerical efficiency (RNE; Geweke, 1992) as well as the convergence of MCMC from different initial conditions. The parameters of interest are, of course, u, χ, ω, λ and Λ.

Geweke's (1992) convergence diagnostic is a z-statistic asymptotically normal (in the number of MCMC draws) that tests for the equality of posterior means in the first 25% and the last 25% of the retained MCMC draws. Geweke's (1992) RNE is a measure of numerical efficiency that should be equal to one if i.i.d. sampling from the posterior were feasible. Autocorrelation functions, convergence diagnostics and RNEs are reported in Figures B.1, B.2, and B.3, respectively. The different "series" correspond to 50 randomly chosen elements of χ, ω, u, as we also do for for ϑ in Figure B.4.

Figure B.1: Autocorrelation functions

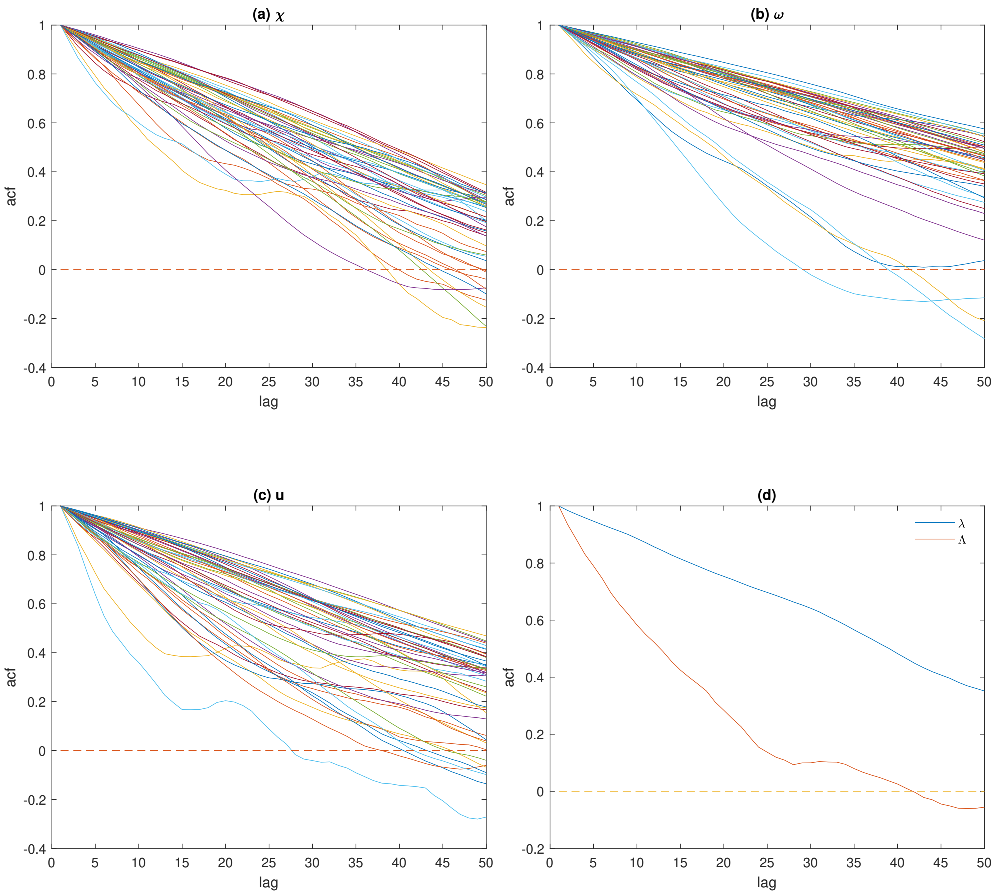


Figure B.2: Geweke (1992) convergence diagnostics

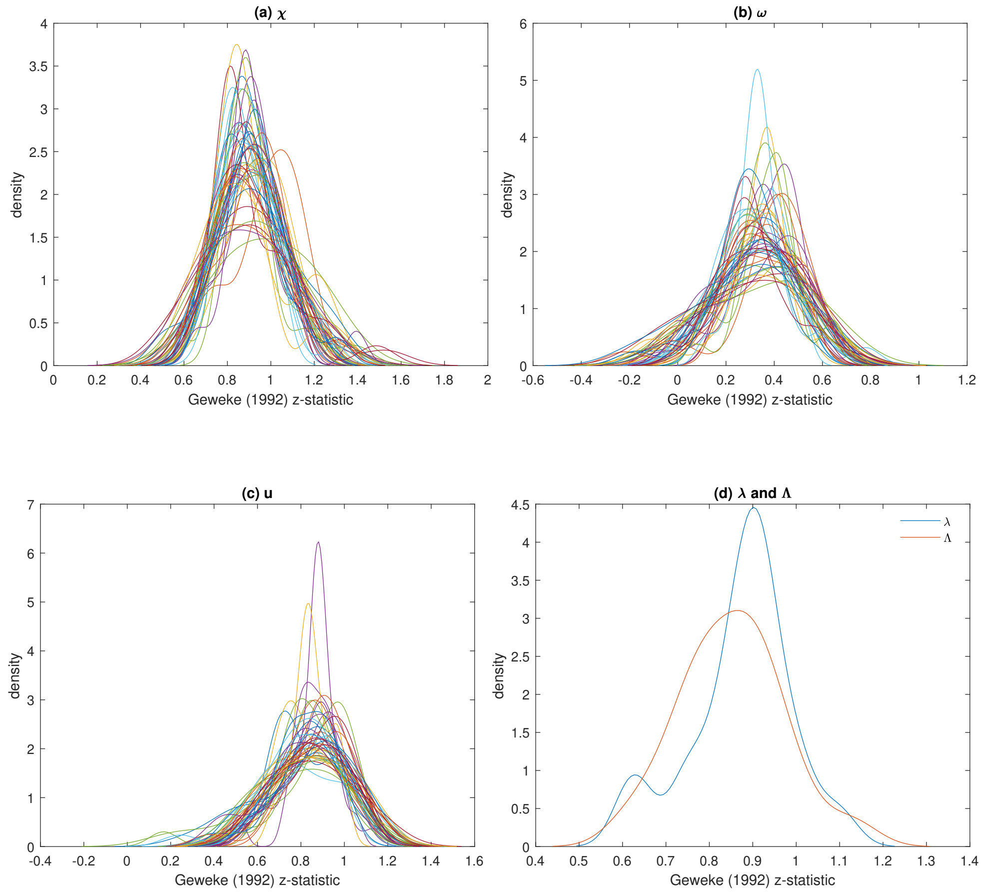
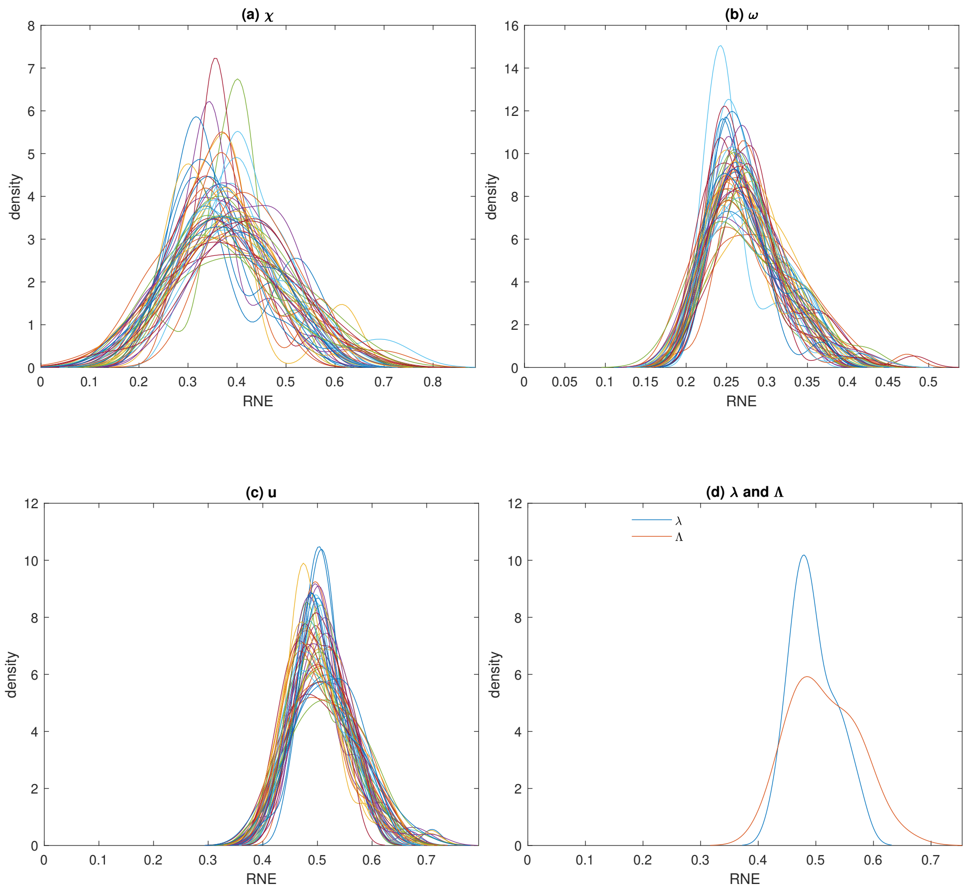
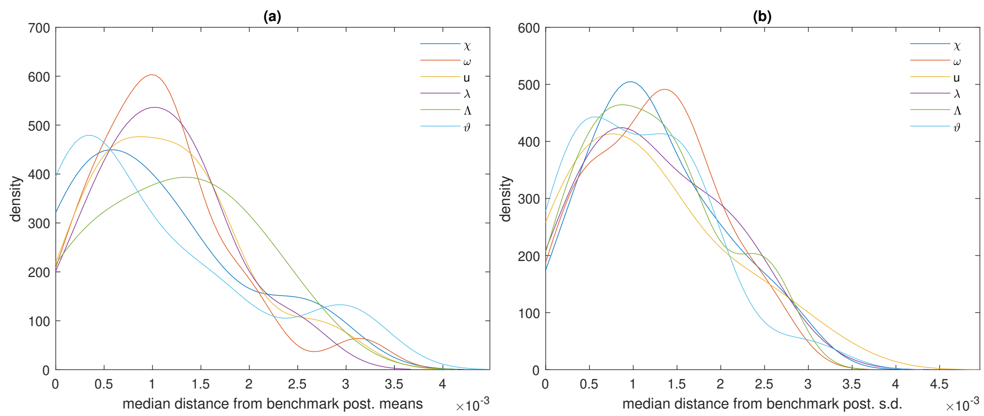


Figure B.3: Relative Numerical Efficiency



In panel (a) of Figure B.4 we present the distributions of distances (L_1-norms) of posterior means computed from MCMC starting from different initial conditions, relative to the benchmark posterior means with the same benchmark prior. In panel (b) of Figure B.4 we do the same with the posterior standard deviations across all parameters and latent variables.

Figure B.4: Convergence from different initial conditions



The evidence shows that MCMC performs well in terms of acf's and RNEs as well as in terms of convergence from different initial conditions. Despite the fact that there is some autocorrelation in MCMC, it is not destructively large so as to prevent thorough exploration of the posterior distributions.