stochastic frontier: determinants of inefficiencies

Ruei-Chi Lee

April 24, 2021

1 Introduction

In this paper, we propose a Bayesian method for four component stochastic frontier panel data models that incorporate persistent (long-term) inefficiency, transient (short-term) inefficiency and firm heterogeneity simultaneously. In particular, we consider factors that can explain persistent and transient inefficiencies. The reason behind that is managers of firms may want to understand not only how many persistent and transient inefficiencies the company made but also what factors caused these inefficiencies, so managers can supervise these factors and improve the efficiencies. Followed from Lai and Kumbhakar, 2018, they consider determinants in the variance parameters of inefficiencies and estimate the model by maximum simulated likelihood (MSL) method. However, if we consider determinants only into the variance, the relationship between inefficiency and determinants will be monotone on average, which is not realistic. For instance, the relationship between financial ratios of firms and efficiencies should be positive while financial ratio is low because it indicates firms improve their turnover rate. On the other hand, if financial ratio is high, the relationship should be negative because firms could not find proper subject to reinvest their fund. Therefore, it is important to consider non-monotonic efficiencies in stochastic frontier model to better understand the relationship between efficiencies and factors. We follow methods from Wang, 2002 to include factors into mean and variance parameters of inefficiencies concurrently.

Four component stochastic fronteir model without determinants was firstly proposed by Colombi et al., 2011. Although they face T dimensional integration from closed skew normal, they use standard maximum likelihood approach to estimate the model. Tsionas and Kumbhakar, 2014 propose a two-step parametrization of Bayesian approach to increase the efficiency of MCMC scheme. Filippini and Greene, 2016 provide MSL method that is comparatively easy to compute the integration from T dimensional to 1. Specifically, since off-diagonal variance are uncorrelated in the closed-skew normal distribution, the likelihood function only require one-dimensional integration. Then they simulate persistent inefficiency and firm heterogeneity from standard normal distribution and calculate the joint pdf. Thus, MSL method have computational advantages in high T cases. As mentioned before, Lai and Kumbhacker (2018) use MSL method to estimate their model that consider factors in variance of inefficiencies due to its computational advantage. However, when considering determinants in both mean and variance, we can not apply MSL directly. Because linear combination is not available when mean parameter of truncated normal is not zero, simulation method from standard normal becomes impractical.

As a result, we propose Bayesian methods to estimate the model. There are several benefits of Bayesian methods for our model compared to ML methods we mentioned above. First, Bayesian methods have good small sample property because we can incorporate prior experience into prior distributions. Second, since we assume random effects and independent between both inefficiencies and firm heterogeneity, we can directly draw these random effects by data augmentation of Bayesian approach. There is no necessary to use simulation method in Lai and Kumbhacker (2018) to acquire inefficiencies and average marginal effects, so we can directly observe inefficiencies and calculate average marginal effects. Third, since Bayesian methods extend parameter space from data augmentation, it don't need to address T dimensional integration from standard ML method. Besides, MSL method is not practical anymore after we consider determinants both in mean and variance parameters of inefficiencies. In this paper, we provide data augmentation, two-step parametrization (Tsiona and Kumbhakar 2014) and Particle Markov chain Monte Carlo (PMCMC) of Bayesian method to estimate the model and compare their results.

In this paper, we show simulation results that PMCMC approach outperform data augmentation and two-step parametrization MCMC. The motivation of providing different Bayesian methods is mixing problem often occurs if number of parameters or random effects are high while using data augmentation method. Tsiona and Kumbhakar (2014) show that parametrization method has faster convergence and lower autocorrelation performance compared to data augmentation in four component stochastic frontier model without determinants case. Thus, we also include and compare their method in this paper. On the other hand, we followed examples from Andrieu et al., 2010 that PMCMC can effectively decrease autocorrelation while drawing these parameters. On the other hand, since particles are independent, we can implement parallel computation for particles to speed up the computation time. Hence, this method can provide better simulation result without sacrificing time complexity. Therefore, we propose a particle Metropolis-within-Gibbs (PMwG) for the model that applies PG (particle gibbs) for random effects and HMC (Hamilton Monte Carlo) for unknown posterior distributions.

2 Model

Consider a four random-component stochastic cost frontier model with inefficient determinants:

$$
\begin{aligned}
y_{it} &= x'_{it}\beta + \alpha_i + \eta_i + u_{it} + v_{it}, \quad i = 1, ..., N \quad t = 1, ..., T \\
\alpha_i &\sim N(0, \sigma^2_\alpha) \\
\eta_i &\sim N^+(w'_i z, e^{w'_i \gamma}) \\
u_{it} &\sim N^+(\pi'_{it}\delta, e^{\pi'_{it}\xi}) \\
v_{it} &\sim N(0, \sigma^2_v)
\end{aligned}
\tag{1}
$$

where y is the dependent variable, x is the independent variable, α_i is the firm heterogeneity effect, v_{it} is a noise term, η_i is the persistent (long-run) inefficiency, and u_{it} is the transient (short-run) inefficiency. w and π are exogenous variables that explain persistent and transient inefficiency separately.

In this model, we assume all random errors, α, η, u, v, are independent. The likelihood of each individual for T observation, marginalized with respect to the random effect α_i, η_i and u_{it} is

$$(L_i|\alpha_i, \eta_{it}, u_{it}) = \prod_{t=1}^{T} \phi(\frac{y_{it} - x'_{it}\beta - \alpha_i - \eta_i - u_{it}}{\sigma_v})$$

$$L_i = \int_0^\infty \int_0^\infty \int_{-\infty}^\infty (L_i|\alpha_i, \eta_{it}, u_{it}) f(\alpha) f(\eta) f(u) d\alpha d\eta du \tag{2}$$

from this we see that the likelihood function for the whole sample is the product of L_i over the N individuals, and the log-likelihood is

$$lnL = \sum_{i=1}^{N} lnL_i$$

The multiple integral does not exist in closed form and the maximization of the log likelihood is complex and time consuming when T is large. Filippini and Greene (2016) adopt the maximum simulated likelihood approach to estimate the four random-components model without determinants of inefficiency. At next section we will propose Bayesian methods to estimate both types of inefficiency and also compute marginal effects of determinants for both persistent and transient technical inefficiencies.

3 Bayesian inference and particle Markov chain Monte Carlo samplers

This section discuss the efficient Bayesian inference for the model we described in Section 2. The first method we applied here is data augmentation of Bayesian approach that is frequently used in missing data or unobserved random variables, such as limited dependent models or panel data models. Since Bayesian inference takes parameters as distributions, data augmentation could treat random variables (α, η, u) as parameters to estimate. As a result, this method extend the original parameters from θ to (θ, α, η, u), so we will not need to handle the multiple integral of likelihood in (2).

Although data augmentation is a flexible method to estimate extended parameters, it still have some problems. Firstly, since the parameters space is extended, it requires large time to reach stationary posterior distributions. Usually, we will use the auto-correlation plot and trace plot to check the mixing problem and stationary status.

To address the mixing problem, two-step parametrization of MCMC scheme is applied to this model. We follow method from Tsiona and Kumbhaker (2014). The first step is parametrized α_i and η_i to τ_i (i.e. $\tau_i = \alpha_i + \eta_i$). The second step is to combine α_i and v_{it} to ψ_{it} (i.e. $\psi_{it} = \alpha_i + v_{it}$). According to this, number of parameters in the MCMC scheme is reduced thus provide a more efficient way to draw. The second method we introduce here is particle Markov chain Monte Carlo (PMCMC) approaches of Andrieu et al. (2010). In the article, their method effectively decrease the auto-correlation if they increase the number of particles, so here I apply their method for drawing random variables effectively.

The PMCMC approach basically define a target distribution on a augmented space that includes parameters and multiple copies (particles) of random effects. Andrieu et al. (2010) provides two ways of PMCMC approaches. First one is the Pseudo Marginal Metropolis-Hasting (PMMH) sampler and the second is the Particle Metropolis within Gibbs (PMwG) sampler. Here we consider PMwG method because the random walk proposal in the PMMH is inefficient from model complexity.

The PMwG method generates the parameters conditioning on the latent random effects and the parameters of the models can be sampled in separate Gibbs or Metropolis within Gibbs step. Besides, by conditioning on the latent states we are interested in, we are able to compute the gradient of log posterior analytically. Thus, for the posterior distributions that requires Metropolis Hasting methods, we apply Hamilton Monte Carlo (HMC) to sample the parameters efficiently.

3.1 PMwG

Here we specify the PMwG method and prior assumptions for the four component stochastic frontier model with determinants of inefficiencies.

Let $\theta = (\beta, \delta, \xi, z, \gamma, \sigma_\alpha^2, \sigma_v^2)$ be the set of target parameters. We use the following prior distributions: $s \sim N(0, \sigma_s^2)$ and $\frac{Q}{\sigma_k^2} \sim \chi^2(\bar{N}_k)$ for $s = \beta, \delta, \xi, z, \gamma$ and $k = \alpha, v$. For hyperparameters we set $Q = 0.0001$, $\sigma_s^2 = 10$ and $\bar{N}_k = 1$

Algorithm 1: PMwG sampling scheme for our model

sample $\beta \sim N((\frac{X'X\sigma_\beta^2+\sigma_v^2}{\sigma_\beta^2})^{-1}X'(y-\alpha-\eta-u), (\frac{X'X\sigma_\beta^2+\sigma_v^2}{\sigma_\beta^2\sigma_v^2})^{-1})$

sample $\sigma_v^{-2} \sim Ga(\frac{NT+\bar{N}_v}{2}, \frac{2}{Q+(y-X\beta-u-\alpha-\eta)'(y-X\beta-u-\alpha-\eta)})$

sample $\sigma_\alpha^{-2} \sim Ga(\frac{N+1}{2}, \frac{2}{Q+\alpha'\alpha})$

sample α, η and u using conditional Monte Carlo method together

sample δ, ξ, z and γ using PMwG with Hamiltonian proposal separately

Algorithm 2: Conditional Monte Carlo for random effects

Fixed $\alpha^1 = \alpha$, $\eta^1 = \eta$, $u^1 = u$.

for $i = 1 : N$ **do**

1. sample M_i^j from $m_i(M_i|\theta, y_i)$[1], for $j = 2, ..., H$, where $M = \alpha, \eta, u$
2. Compute the importance weight: $w_i^j = \frac{p(y_i|\alpha_i^j,\theta,\eta_i,u_i)p(\alpha_i^j|\theta)p(\eta_i^j|\theta)p(u_i^j|\theta)}{m_i(\alpha_i|\theta,y_i)m_i(\eta_i|\theta,y_i)m_i(u_i|\theta,y_i)} = p(y_i|\alpha_i^j, \theta, \eta_i^j, u_i^j)$
3. Normalised the weight $\bar{w}_i^j = \frac{w_i^j}{\sum_{k=1}^H w_i^k}$
4. sample index k_i from $Pr(k_i) = \bar{w}_i^j$ and decide $\alpha_i^{k_i}, \eta_i^{k_i}$ and $u_i^{k_i}$

Algorithm 3: Data augmentation for random effects

sample $\eta \sim N^+((\Sigma_\eta^{-1} + \frac{T}{\sigma_v^2}I_N)^{-1}(\frac{T}{\sigma_v^2}I_N\frac{1}{T}\sum_t(y - X\beta - \eta - \alpha - u) + \Sigma_\eta^{-1}wz), (\Sigma_\eta^{-1} + \frac{T}{\sigma_v^2}I_N)^{-1})$

sample $u \sim N^+((\frac{I_{NT}}{\sigma_v^2} + \Sigma_u^{-1})^{-1}(\frac{y-X\beta-\eta-\alpha}{\sigma_v^2} + \Sigma_u^{-1}\pi\delta), (\frac{I_{NT}}{\sigma_v^2} + \Sigma_u^{-1})^{-1})$

sample $\delta \propto N((\frac{I_{NT}}{\sigma_\delta^2} + \pi'\Sigma_u^{-1}\pi)^{-1}\pi'\Sigma_u^{-1}u, (\frac{I_{NT}}{\sigma_\delta^2} + \pi'\Sigma_u^{-1}\pi)^{-1})\Phi^{-1}(\frac{\pi\delta}{exp\frac{\pi\xi}{2}})$

[1] $m_i(\alpha_i|\theta, y_i) = N(0, \sigma_\alpha^2)$, $m_i(\eta_i|\theta, y_i) = N^+(w_i'z, e^{w_i'\gamma})$, $m_{it}(u_{it}|\theta, y_{it}) = N^+(u_{it}|\pi_{it}'\delta, e^{\pi_{it}\xi})$

Algorithm 4: Two-step parametrization MCMC for our model

τ parametrization :
sample $\tau \propto exp\{\frac{-1}{2}\frac{(\tau-\mu_\eta)^2}{\sigma_\alpha^2+\sigma_\eta^2} + \frac{-1}{2}\frac{\sum_t R_{it}^2}{\sigma_v^2}\}\Phi(\frac{\sigma_\eta}{\sigma_\alpha}\frac{\tau\mu_\eta}{\sqrt{{\sigma_\alpha}^2+\sigma_\eta^2}} + \frac{\mu_\eta\sqrt{\sigma_\alpha^2+\sigma_\eta^2}}{\sigma_\alpha\sigma_\eta})$
sample $\beta \sim N((\frac{X'X\sigma_\beta^2+\sigma_v^2}{\sigma_\beta^2})^{-1}X'(y-\tau-\eta), (\frac{X'X\sigma_\beta^2+\sigma_v^2}{\sigma_\beta^2\sigma_v^2})^{-1})$
sample $\sigma_v^{-2} \sim Ga(\frac{NT+\bar{N}_v}{2}, \frac{2}{Q+(y-X\beta-u-\tau)'(y-X\beta-u-\tau)})$
ψ parametrization :
sample $\eta \sim N^+(\varphi^2(\iota_t'\Sigma^{-1}(y-x\beta-u)+\frac{\mu_\eta}{\sigma_\eta^2}), \varphi^2)$, where
$\varphi = (\iota_T'\Sigma^{-1}\iota_T + \sigma_\eta^{-2})^{-1}$, $\Sigma = \sigma_\alpha^2 J_T + \sigma_v^2 I_T$
sample $\alpha = \tau - \eta$
sample $\sigma_\alpha^{-2} \sim Ga(\frac{N+\bar{N}_\alpha}{2}, \frac{2}{Q+\alpha'\alpha})$
sample δ, ξ, z and γ using MCMC with Hamiltonian proposal separately

In order to compare the efficiency of methods, here we can replace the step 3 of algorithm 1 from Conditional Monte Carlo to data augmentation. Algorithm 4 is the two-step parametrization method that designed for the model. The next section will mainly compare their simulation results.

4 Simulation

In this section, we provide our DGA and prior setting for our model. Here we set number of individuals $N = \{100, 500, 1000\}$, time period $T = 10$, number of particles $H = 50000$. The true parameter are set as $\sigma_\alpha^2 = 0.25$, $\sigma_v^2 = 0.25$ and rest parameters $\beta_0 = \beta_1 = \delta_0 = \delta_1 = \xi_0 = \xi_1 = z_0 = z_1 = \gamma_0 = \gamma_1 = 1$. The independent variables x, w, π are drawn from standard normal. We draw 11000 samplers and discard first 1000 as burn-in.

For comparison, data augmentation, two-step parametrization MCMC and PMwG estimation are proposed and summarized their bias and mean square error (MSE) in Table 1, 2 and 3, respectively. These tables indicate that bias and MSE of most parameters become smaller as sample size increase. Specifically, Bias and MSE of data augmentation of σ_α^2 does not improve even though we increase the sample size from $N = 100$ to 1000. The estimation of σ_α^2 remain negligible and so does firm heterogeneity α_i. In other words, when we implement data augmentation method, firm heterogeneity is absorbed due to multiple random effects this method need to draw. For two-step estimator, Bias and MSE of σ_α^2 and σ_v^2 don't ameliorate despite increasing the sample size. The reason comes from the τ parametrization step because firm heterogeneity α is decided after τ and δ are drawn. This arbitrary step makes α can not adjust itself and the σ_α^2 stays higher than 0.25 we set.

On the other hand, Conditional Monte Carlo of PMwG method (Algorithm 2) for random effects are proposed. H set of random effects candidates are drawn and one of them is selected by importance weight. The results of Bias and MSE of σ_α^2 are highly improved compared to other methods. It shows that PMwG has better performance for latent variables than data augmentation method.

Finally, in order to examine the trade-off between computation time and number of particles of PMwG, we plot autocorrelation (Figure 4) and computation time in different setting of number of particles of PMwG. Since more particles have preciser likelihood estimation with more computational burden, Figure 4 shows PMwG provides lower autocorrelation of parameters. On the other hand, smaller particles have less computational cost (Figure 5) but slow mixing property in most cases. In summary, PMwG with higher number of particles could provide better performance in autocorrelation, so PMwG is indeed improved mixing property as number of particles increases.

Parameter	Bias			MSE		
	N=100	N=500	N=1000	N=100	N=500	N=1000
$\beta_0 = 1$	-0.6896	-0.4764	-0.4151	0.5967	0.2345	0.1777
$\beta_1 = 1$	0.0198	-0.0145	-0.0046	0.0015	0.0004	0.0001
$\delta_0 = 1$	-0.531	-0.086	-0.0428	0.5293	0.0325	0.0123
$\delta_1 = 1$	0.0528	0.0588	0.0143	0.078	0.0183	0.0062
$\xi_0 = 1$	0.0649	0.0457	-0.0095	0.0425	0.0076	0.0027
$\xi_1 = 1$	-0.0982	-0.0366	-0.0311	0.0272	0.0052	0.0029
$z_0 = 1$	1.2481	0.7994	0.8486	1.7485	0.6996	0.7439
$z_1 = 1$	-0.2149	-0.0544	-0.1607	0.0656	0.0952	0.0435
$\gamma_0 = 1$	-0.782	0.0787	-0.1023	0.6756	0.0267	0.0198
$\gamma_1 = 1$	0.4932	-0.3322	-0.1309	0.2757	0.1477	0.0294
$\sigma_\alpha^2 = 0.25$	-0.25	-0.25	-0.25	0.0625	0.0625	0.0625
$\sigma_v^2 = 0.25$	0.0624	-0.0028	0.0324	0.0078	0.0005	0.0014

Table 1: Bias and MSE of data augmentation estimator

Parameter	Bias			MSE		
	N=100	N=500	N=1000	N=100	N=500	N=1000
$\beta_0 = 1$	0.2455	0.2605	0.407	0.1093	0.08	0.173
$\beta_1 = 1$	-0.0307	0.0289	-0.001	0.0021	0.0011	0.0001
$\delta_0 = 1$	0.373	-0.2887	0.0636	0.2079	0.1193	0.0135
$\delta_1 = 1$	-0.1324	0.0482	-0.0513	0.0537	0.022	0.0072
$\xi_0 = 1$	-0.1133	0.1078	-0.1161	0.031	0.0184	0.0163
$\xi_1 = 1$	0.0601	-0.0243	0.0654	0.018	0.0053	0.0062
$z_0 = 1$	-1.0587	-0.8138	-1.6659	3.4014	1.1962	3.8834
$z_1 = 1$	0.6757	0.3103	0.6691	1.3458	0.2303	0.9752
$\gamma_0 = 1$	-0.0312	-0.1815	0.1605	0.4228	0.1359	0.1327
$\gamma_1 = 1$	0.0169	0.0198	0.0918	0.1682	0.0365	0.0596
$\sigma_\alpha^2 = 0.25$	0.2755	0.3182	0.4099	0.1151	0.1078	0.1722
$\sigma_v^2 = 0.25$	-0.0397	0.0218	0.0356	0.0045	0.0011	0.0016

Table 2: Bias and MSE of Two-step estimator

Parameter	Bias			MSE		
	N=100	N=500	N=1000	N=100	N=500	N=1000
$\beta_0 = 1$	-0.0682	0.0705	-0.0245	0.0124	0.0092	0.0018
$\beta_1 = 1$	0.0477	-0.032	-0.0043	0.0031	0.0012	0.0001
$\delta_0 = 1$	-0.5082	0.3994	0.0407	0.4163	0.1687	0.0097
$\delta_1 = 1$	-0.0329	-0.0691	-0.0452	0.0745	0.01	0.0081
$\xi_0 = 1$	0.2565	-0.1745	0.0184	0.0906	0.034	0.0029
$\xi_1 = 1$	-0.0301	0.0955	-0.0125	0.0129	0.012	0.0022
$z_0 = 1$	-0.8573	-0.3475	-0.0969	2.4192	0.3415	0.1029
$z_1 = 1$	0.192	0.0617	0.254	0.7791	0.0787	0.1676
$\gamma_0 = 1$	0.3985	-0.014	0.1772	0.3445	0.0542	0.0563
$\gamma_1 = 1$	-0.1823	-0.0866	-0.1022	0.1432	0.0316	0.0329
$\sigma_\alpha^2 = 0.25$	-0.1346	0.0587	0.0072	0.0227	0.007	0.0013
$\sigma_v^2 = 0.25$	-0.0253	-0.0102	-0.0039	0.0017	0.0003	0.0003

Table 3: Bias and MSE of PMwG estimator

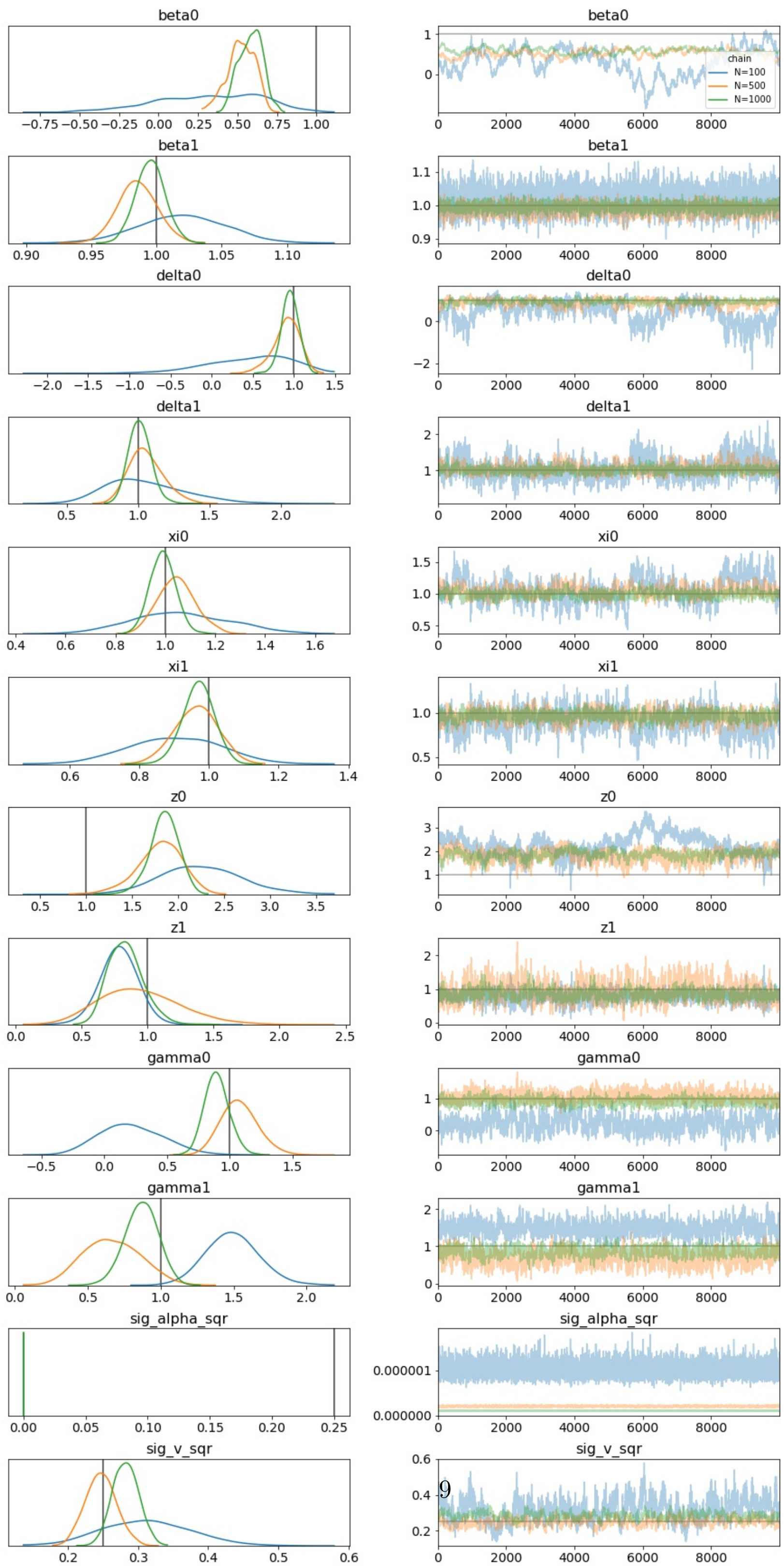

Figure 1: trace plot of data augmentation

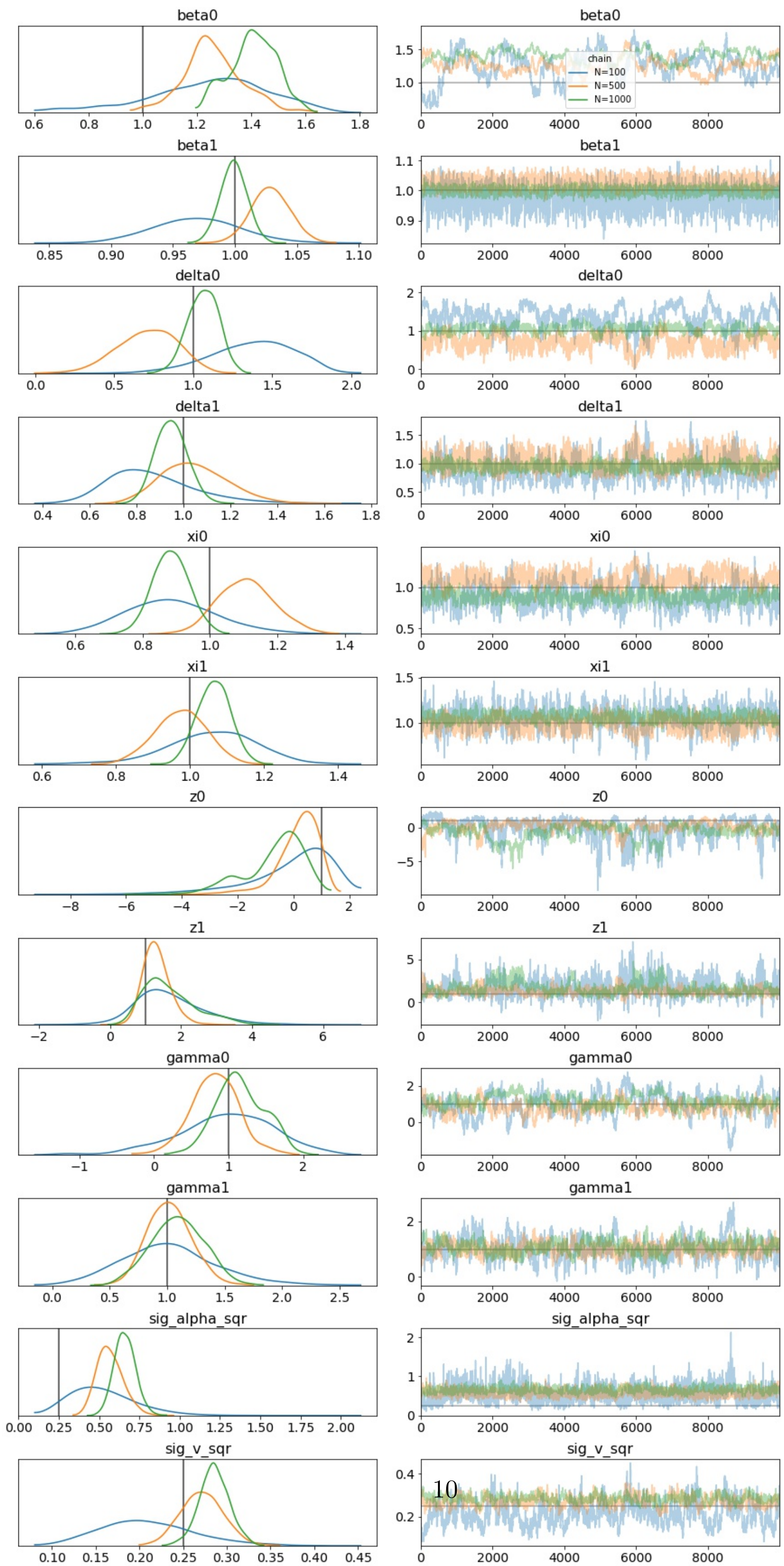

Figure 2: trace plot of two-step parametrization

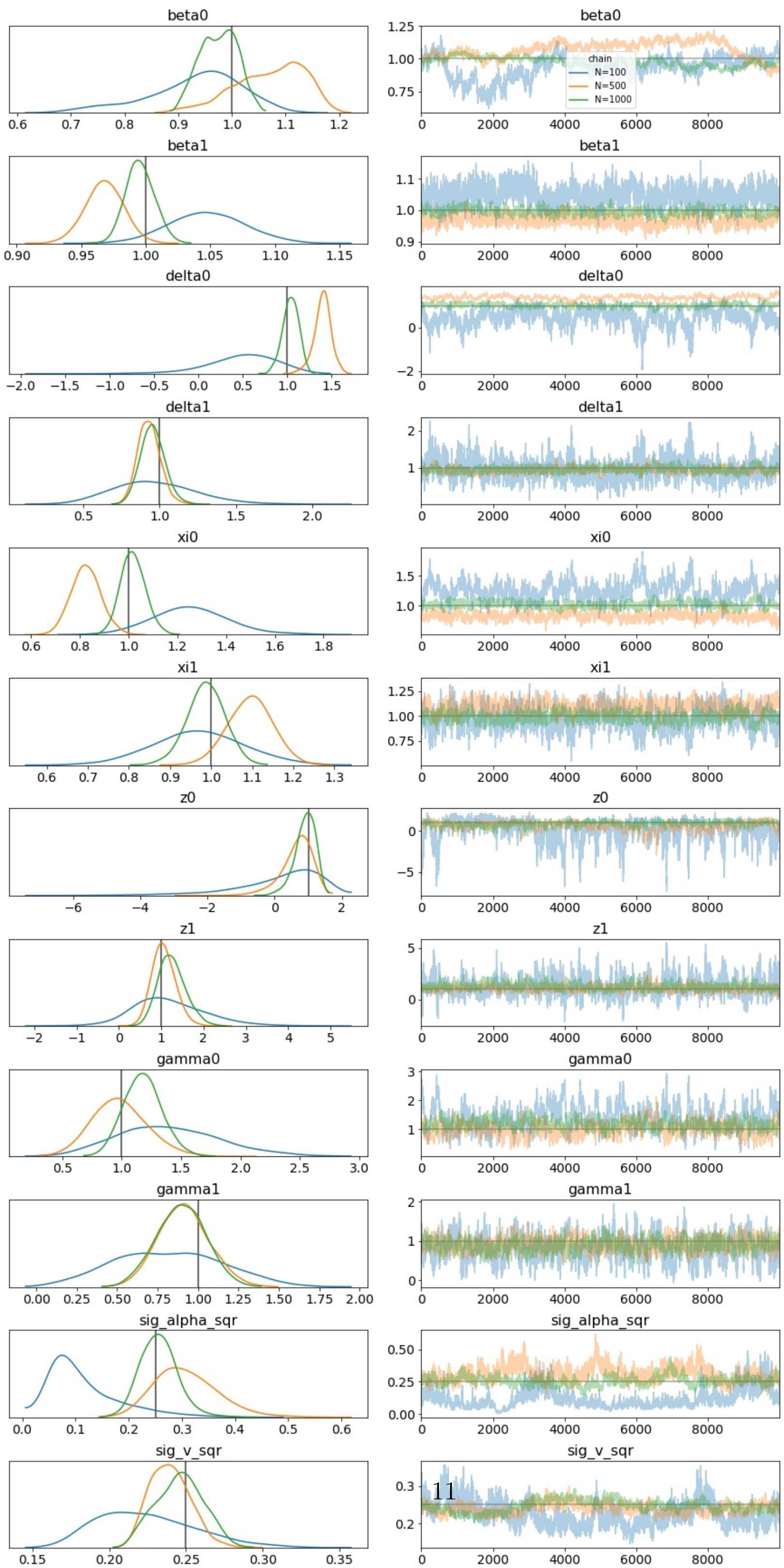

Figure 3: trace plot of PMwG

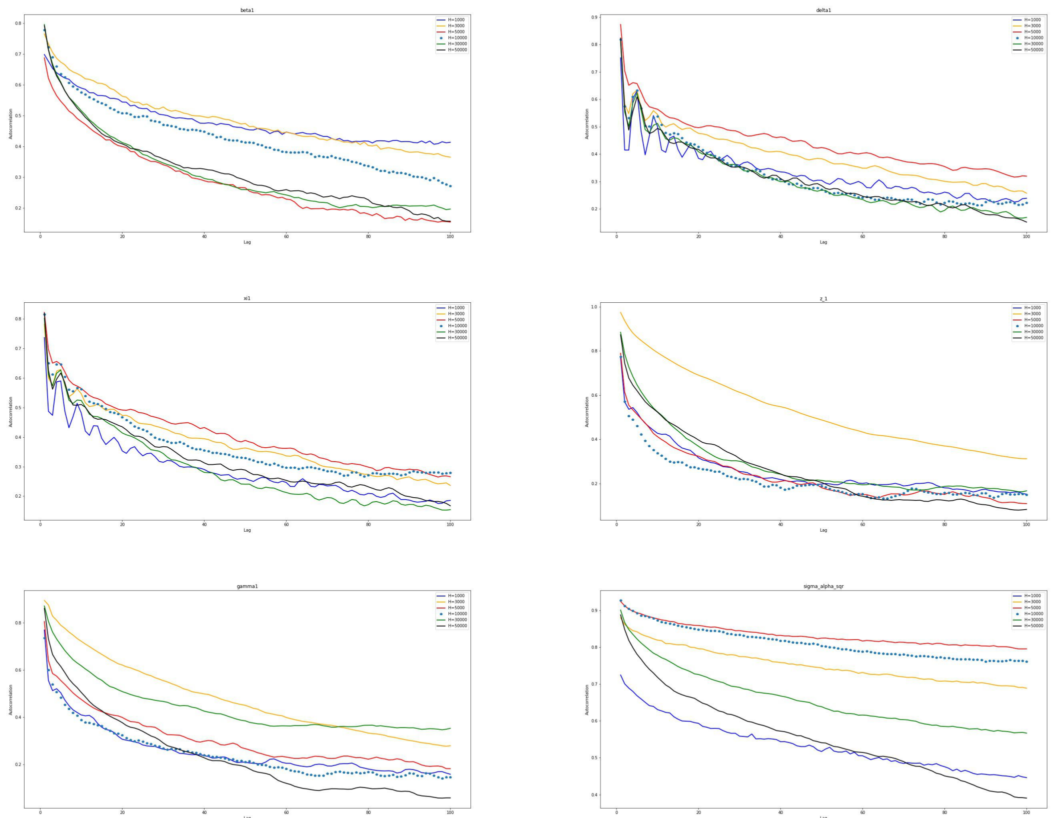

Figure 4: autocorrelation of PMwG with different H

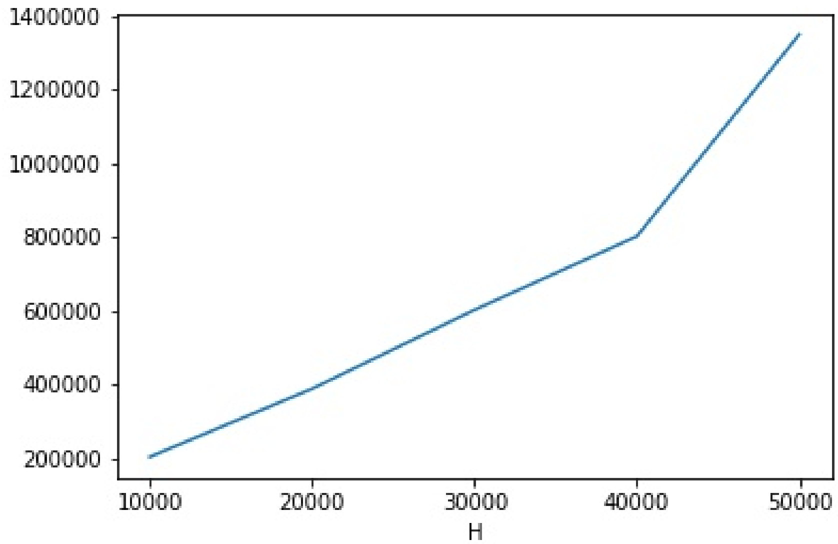

Figure 5: computation time (seconds) of PMwG with different H

References

Andrieu, C., Doucet, A., & Holenstein, R. (2010). Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *72*(3), 269–342. https://doi.org/https://doi.org/10.1111/j.1467-9868.2009.00736.x

Colombi, R., Martini, G., & Vittadini, G. (2011). *A stochastic frontier model with short-run and long-run inefficiency random effects* (Working Papers No. 1101). Department of Economics and Technology Management, University of Bergamo. https://EconPapers.repec.org/RePEc:brh:wpaper:1101

Filippini, M., & Greene, W. (2016). Persistent and transient productive inefficiency: A maximum simulated likelihood approach. *Journal of Productivity Analysis*, *45*(2), 187–196. https://EconPapers.repec.org/RePEc:kap:jproda:v:45:y:2016:i:2:p:187-196

Lai, H.-p., & Kumbhakar, S. C. (2018). Panel data stochastic frontier model with determinants of persistent and transient inefficiency. *European Journal of Operational Research*, *271*(2), 746–755. https://doi.org/https://doi.org/10.1016/j.ejor.2018.04.043

Tsionas, E. G., & Kumbhakar, S. C. (2014). Firm heterogeneity, persistent and transient technical inefficiency: A generalized true random-effects model. *Journal of Applied Econometrics*, *29*(1), 110–132. https://doi.org/https://doi.org/10.1002/jae.2300

Wang, H.-J. (2002). Heteroscedasticity and non-monotonic efficiency effects of a stochastic frontier model. *Journal of Productivity Analysis*, *18*(3), 241–253. https://EconPapers.repec.org/RePEc:kap:jproda:v:18:y:2002:i:3:p:241-253