Bayesian inference in regression models with the restricted parameter spaces

Hossein Bevrani[1], Solmaz Seifollahi[1], Kaniav Kamary[2]

[1]Department of Statistics, University of Tabriz

[2]CentraleSupélec, Université Paris-Saclay, Gif-sur-Yvette, France

September 8, 2023

Tabriz, Iran
(work place)


Bayesian inference in regression models with the restricted parameter space

Recently, the subject that raised commonly in the framework of regression models, especially in the linear regression models and also generalized linear models, is estimating model parameters based on uncertain prior information. Taking these uncertain prior information into account when estimating the model can improve the estimation procedure in terms of accuracy, and so incorporating them has recently attached a lot of research attention. The uncertain prior information is incorporated in the model through some restrictions on parameters.

Note

The restrictions can be linear or nonlinear functions of parameters, or even in the form of equality or inequality.

Linear equality restrictions

Where these restrictions come from?

- Obtained according to past data or expert opinion.
- From variable selection techniques such as AIC, CAIC and BIC.

Consider the linear equality restrictions on the model parameters that can be written as:

$$\boldsymbol{R\beta} = r$$

where $R_{q\times p}$ and $r_{q\times 1}$ are a pre-specified matrix and vector.

Linear equality restrictions

What are the estimators?

- Unrestricted estimator
- Rerestricted estimator
- Shrinkage methods
  1. Linear shrinkage estimator
  2. Pretest estimator
  3. James-Stein type estimator

Linear inequality restrictions

Where these kind of restrictions come from?

In the real world, physical quantities show specific behaviors:

- Chemical concentrations lies between 0 and 1.
- Due to the physical nature, Some models usually have restrictions on the sign of the input variables
- In econometrics, we encounter situations where some parameters of the model should be considered non-negative or negative.
- In hyperspectral imaging, due to physical considerations, coefficient parameters should be considered non-negative.

What to do?

From a partical point of view, it would be highly desirable to build a surragate model which respects the same restrictionsto ensure structural compatibility based on physical phenomena or the validity of scientific theories.

The linear inequality restrictions on the model:

$$L_{q\times 1} \leq H_{q\times p}\beta_{p\times 1} \leq G_{q\times 1}$$

It contains:

- linear equality restrictions
- Order restrictions
- Sign restrictions

Bayesian inference in Linear regression model with linear inequality restrictions

The linear regression model is given by:

$$Y = \boldsymbol{X\beta} + \varepsilon, \qquad \varepsilon \sim N(0, \sigma^2)$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_p)^T$ is a vector of the regression coefficients. Consider q linear inequality restrictions on the model parameters as:

$$K \leq H\boldsymbol{\beta} \leq G$$

where H is a matrix of size $q \times p$ with $(q \leq p)$ and G and K are two vectors of length q.

Classical inference

- Judge and Takayama (1966) used Dantzig-Cottle algorithm (IROLS)
- Escobar and Skarpness (1986, 1987) obtained the bias and efficiency of IROLS
- Ohtani (1987) obtained the MSE of IROLS

Bayesian inference

- Geweke (1986):
  - $\pi(\boldsymbol{\beta}, \sigma^2) \propto \frac{1}{\sigma} I_{(K \leq H\boldsymbol{\beta} \leq G)}$
  - Method: importance sampler
  - When p increases or the posterior probability of restriction is small, the algorithm will be slow.
- Geweke (1996)
  - Method: Gibbes sampler
  - Algorithm is practical when H is invertable.
- Rodriguez-Yam et al. (2004):
  - Method: Gibbes sampler
  - Algorithm is practical when the coefficients only are upper bounded.

Prior specification

We Partition $H = (H_S, H_{S'})$

How?

Suppose that $P = \{1, \ldots, p\} = S \cup S'$ is a set of the indices of the matrix H.
$S = \{j : h_{ij} \in H_S; i = 1, \ldots, q\} \subset P$ is a subset of the indices of the columns of H for which the block H_S is a full rank matrix.

The result

The matrix H_S is then a $q \times q$ full rank block of H.
$H_{S'}$ is a $q \times (p - q)$ sub-matrix including the rest of the elements of H.

Consequence

$$\boldsymbol{\beta} = (\boldsymbol{\beta}_S^T, \boldsymbol{\beta}_{S'}^T)^T$$

Therefore:

$$K \leq \left(H_{S(q\times q)}, H_{S'(q\times(p-q))}\right) \begin{pmatrix} \boldsymbol{\beta}_{S(q\times 1)} \\ \boldsymbol{\beta}_{S'((p-q)\times 1)} \end{pmatrix} \leq G$$

$$K \leq H_S \boldsymbol{\beta}_S + H_{S'} \boldsymbol{\beta}_{S'} \leq G$$

$$K - H_{S'} \boldsymbol{\beta}_{S'} \leq H_S \boldsymbol{\beta}_S \leq G - H_{S'} \boldsymbol{\beta}_{S'}$$

Prior distributions

1. $\sigma^2 \sim IG(\frac{a}{2}, \frac{b}{2})$
2. $\boldsymbol{\beta}_{S'}|\sigma^2 \sim N(\mu_{\boldsymbol{\beta}_{S'}}, \sigma^2\Sigma_{\boldsymbol{\beta}_{S'}})$
3. $\boldsymbol{\beta}_S|\boldsymbol{\beta}_{S'}, \sigma^2 \sim TN(\mu_{\boldsymbol{\beta}_S}, \sigma^2\Sigma_{\boldsymbol{\beta}_S}, T)$ where

$$T = \{\boldsymbol{\beta}_S \in \mathbb{R}^q | K - H_{S'}\boldsymbol{\beta}_{S'} \leq H_S\boldsymbol{\beta}_S \leq G - H_{S'}\boldsymbol{\beta}_{S'}\}.$$

Posterior distributions

$$Y = \boldsymbol{X}_{S;S'}\boldsymbol{\beta}_{S;S'} + \boldsymbol{\varepsilon}$$
$$= (\boldsymbol{X}_S, \boldsymbol{X}_{S'}) \begin{pmatrix} \boldsymbol{\beta}_S \\ \boldsymbol{\beta}_{S'} \end{pmatrix} + \boldsymbol{\varepsilon}$$
$$= \boldsymbol{X}_S\boldsymbol{\beta}_S + \boldsymbol{X}_{S'}\boldsymbol{\beta}_{S'} + \boldsymbol{\varepsilon}$$

The posterior distributions will be:

- $\sigma^2 | Y, \boldsymbol{X}_{S;S'} \sim IG(\tilde{v}, \tilde{\eta})$
  where

$$\tilde{v} = \frac{n+a}{2}$$
$$\tilde{\eta} = \frac{1}{2}(b + Y^T Y + \mu_{\beta_{S'}}^T \Sigma_{\beta_{S'}}^{-1} \mu_{\beta_{S'}} + \mu_{\beta_S}^T \Sigma_{\beta_S}^{-1} \mu_{\beta_S} - W^T \tilde{\Sigma}_{\beta_S} W - \tilde{\mu}_{\beta_{S'}} \tilde{\Sigma}_{\beta_{S'}}^{-1} \tilde{\mu}_{\beta_{S'}})$$

Posterior distributions

- $\boldsymbol{\beta}_{S'}|\sigma^2, Y, \boldsymbol{X}_{S;S'} \sim N(\tilde{\mu}_{\beta_{S'}}, \sigma^2\tilde{\Sigma}_{\beta_{S'}})$ where

$$\tilde{\Sigma}_{\beta_{S'}} = \boldsymbol{X}_{S'}^T\boldsymbol{X}_{S'} + \Sigma_{\beta_{S'}}^{-1} - (\boldsymbol{X}_S^T\boldsymbol{X}_{S'})^T\tilde{\Sigma}_{\beta_S}(\boldsymbol{X}_S^T\boldsymbol{X}_{S'})$$

$$\tilde{\mu}_{\beta_{S'}} = \tilde{\Sigma}_{\beta_{S'}}(\boldsymbol{X}_{S'}^T Y + \Sigma_{\beta_{S'}}^{-1}\mu_{\beta_{S'}} - (\boldsymbol{X}_S^T\boldsymbol{X}_{S'})^T\tilde{\Sigma}_{\beta_S}W)$$

- $\boldsymbol{\beta}_S|\boldsymbol{\beta}_{S'}, \sigma^2, Y, \boldsymbol{X}_{S;S'} \sim TN(\tilde{\mu}_{\beta_S}, \sigma^2\tilde{\Sigma}_{\beta_S}, T)$ where

$$\tilde{\Sigma}_{\beta_S} = \boldsymbol{X}_S^T\boldsymbol{X}_S + \Sigma_{\beta_S}^{-1}$$

$$\tilde{\mu}_{\beta_S} = \tilde{\Sigma}_{\beta_S}(W - \boldsymbol{X}_S^T\boldsymbol{X}_{S'}\boldsymbol{\beta}_{S'})$$

$$W = \boldsymbol{X}_S^T Y + \Sigma_{\beta_S}^{-1}\mu_{\beta_S}$$

Proposed algorithm (Bevrani Kamary Seifollahi (BKS) algorithm)

- Initialize $\boldsymbol{\beta}^{(0)}$ and $\sigma^{2(0)}$,
- Update $\boldsymbol{\beta}^{(t)}$ and $\sigma^{2(t)}$ for $t : 1, 2, \ldots$ by the below steps:
  - Step 1: Draw $\sigma^{2(t)} | Y, \boldsymbol{X}_{S;S'} \sim IG(\tilde{\nu}, \tilde{\eta})$,
  - Step 2: Draw $\boldsymbol{\beta}_{S'}^{(t)} | \sigma^{2(t)}, Y, \boldsymbol{X}_{S;S'} \sim N(\tilde{\mu}_{\boldsymbol{\beta}_{S'}}, \sigma^{2(t)} \tilde{\Sigma}_{\boldsymbol{\beta}_{S'}})$,
  - Step 3: Calculate the restriction inequality bounds and the posterior mean of $\boldsymbol{\beta}_S$ from the following equation:

$$T^{(t)} = \{\boldsymbol{\beta}_S \in \mathbb{R}^q | K - H_{S'} \boldsymbol{\beta}_{S'}^{(t)} \leq H_S \boldsymbol{\beta}_S \leq G - H_{S'} \boldsymbol{\beta}_{S'}^{(t)}\},$$

$$\tilde{\mu}_{\boldsymbol{\beta}_S} = \tilde{\Sigma}_{\boldsymbol{\beta}_S} (W - \boldsymbol{X}_S^T \boldsymbol{X}_{S'} \boldsymbol{\beta}_{S'}^{(t)})$$

  - Step 4: Draw $\boldsymbol{\beta}_S^{(t)} | \boldsymbol{\beta}_{S'}^{(t)}, \sigma^{2(t)}, Y, \boldsymbol{X}_{S;S'} \sim TN(\tilde{\mu}_{\boldsymbol{\beta}_S}, \sigma^{2(t)} \tilde{\Sigma}_{\boldsymbol{\beta}_S}, T^{(t)})$.

Consider $y_i = \beta_1 + \beta_2 x_{i1} + \beta_3 x_{i2} + \beta_4 x_{i3} + \beta_5 x_{i4} + \varepsilon_i \qquad i : 1, 2, \cdots, 20$

$$\varepsilon_i \sim N(0,1)$$
$$x_{ij} \sim N(0,1)$$
$$\boldsymbol{\beta} = (-0.5, 1, -2, 3, 4)^T$$

Restriction 1	Restriction 2
$\beta_2 + \beta_3 \leq -0.5$ $\beta_2 + \beta_4 - \beta_5 \leq 0.2$ $\beta_3 + \beta_5 \leq 2.2$	$\beta_2 + \beta_3 \leq -0.5$ $\beta_3 \leq -1.5$ $\beta_4 \geq 2$
BKS	BKS and Geweke

$$H^{[1]} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix}, \quad G^{[1]} = \begin{pmatrix} -0.5 \\ 0.2 \\ 2.2 \end{pmatrix}$$

$$H^{[2]} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \end{pmatrix}, \quad G^{[2]} = \begin{pmatrix} -0.5 \\ -1.5 \\ -2 \end{pmatrix}$$

$$H_G = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad G_G = \begin{pmatrix} +\infty \\ -0.5 \\ -1.5 \\ -2 \\ +\infty \end{pmatrix}$$

$$H_S^{[1]} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 1 & 0 & 1 \end{pmatrix},\ H_{S'}^{[1]} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 0 \end{pmatrix},\ H_S^{[2]} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix},\ H_{S'}^{[2]} = \mathbf{0}_{3\times 2}$$

$$\boldsymbol{\beta}_S^{[1]} = (\beta_3, \beta_4, \beta_5)^T,\ \boldsymbol{\beta}_{S'}^{[1]} = (\beta_1, \beta_2)^T,\ \boldsymbol{\beta}_S^{[2]} = (\beta_2, \beta_3, \beta_4)^T,\ \boldsymbol{\beta}_{S'}^{[2]} = (\beta_1, \beta_5)^T$$

$$\boldsymbol{X}_S^{[1]} = (X_2, X_3, X_4),\ \boldsymbol{X}_{S'}^{[1]} = (\mathbf{1}_n, X_1),\ \boldsymbol{X}_S^{[2]} = (X_1, X_2, X_3),\ \boldsymbol{X}_{S'}^{[2]} = (\mathbf{1}_n, X_4)$$

Prior distributions:

$$\sigma^2 \sim IG(3,1)$$
$$\boldsymbol{\beta}_{S'}|\sigma^2 \sim N_2(\mu_{S'}, \sigma^2(\boldsymbol{X}_{S'}^T\boldsymbol{X}_{S'})^{-1})$$
$$\boldsymbol{\beta}_S|\boldsymbol{\beta}_{S'}, \sigma^2 \sim TN_3(\mu_S, \sigma^2(\boldsymbol{X}_S^T\boldsymbol{X}_S)^{-1}, M)$$

where

$$\hat{\boldsymbol{\beta}}_{OLS} = \mu = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T Y$$
$$M = \{\boldsymbol{\beta}_S \in \mathbb{R}^3 | H_S\boldsymbol{\beta}_S \leq G - H_{S'}\boldsymbol{\beta}_{S'}\}.$$

Prior distribution for Geweke method:

$$\boldsymbol{\beta}|\sigma^2 \sim N_5(\mu, \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}).I_{(H_G\boldsymbol{\beta} \leq G_G)}$$

First, we generate 10^4 data from the posterior distribution by the proposed algorithm.

	Parameters	stationary test	p-value	Halfwidth test	Mean	Halfwidth
	β_1	passed	0.356	passed	-0.329	0.002
	β_2	passed	0.568	passed	1.195	0.002
Rest 1.	β_3	passed	0.543	passed	-1.996	0.007
	β_4	passed	0.949	passed	2.794	0.002
	β_5	passed	0.260	passed	3.931	0.003
	β_1	passed	0.293	passed	-0.329	0.002
	β_2	passed	0.210	passed	1.192	0.001
Rest 2.	β_3	passed	0.822	passed	-1.994	0.005
	β_4	passed	0.126	passed	2.910	0.002
	β_5	passed	0.550	passed	3.820	0.002

Table : The Heidelberg and Welch diagnostic results of coefficients based on generated data.

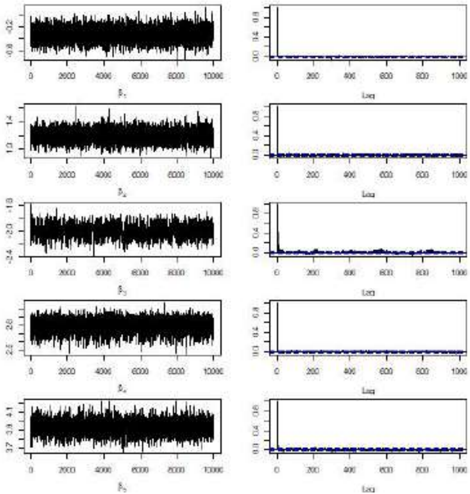

Figure : Trace and the sample ACF plots of generated data in **Rest 1**.

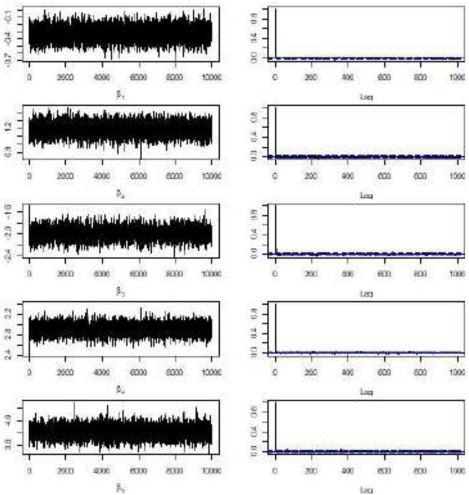

Figure : Trace and the sample ACF plots of generated data in **Rest 2**.

Gibbs path of length 10^4 with 2000 data as burn-in and 2000 replication. The loss function is quadratic function. The criteria for comparing estimators:

$$\mathrm{MSE}(\hat{\boldsymbol{\beta}}) = \frac{1}{2000}\sum_{k=1}^{2000}\sum_{j=1}^{p}(\hat{\beta}_{jk} - \beta_j^{real})^2$$

$$\mathrm{RE} = \frac{\mathrm{MSE}(\hat{\boldsymbol{\beta}}_{UN})}{\mathrm{MSE}(\hat{\boldsymbol{\beta}}_{RE})}$$

			Restriction 1.	Restriction 2.	
		Unrestricted	BKS-method	BKS-method	Geweke's method
β_1	$\hat{\beta}_1$	-0.5019	-0.5016	-0.4984	-0.4997
	SE	0.2610	0.2545	0.2527	0.2556
β_2	$\hat{\beta}_2$	0.9960	1.0013	0.9853	0.9720
	SE	0.2678	0.2614	0.2387	0.2396
β_3	$\hat{\beta}_3$	-1.9990	-2.0002	-2.0301	-2.0342
	SE	0.2718	0.2019	0.2205	0.2270
β_4	$\hat{\beta}_4$	3.0031	2.9983	3.0018	2.9955
	SE	0.2655	0.2216	0.2630	0.2637
β_5	$\hat{\beta}_5$	3.9957	4.0025	4.0009	4.0002
	SE	0.2793	0.1831	0.2580	0.2608
MSE		0.0724	0.0557	0.0602	0.0627
RE		-	1.30	1.20	1.15
time		-	-	66.54	82.64

Table : Bayesian estimations and SE of parameters with different methods

We considered the illustration provided by Pindyck and Rubinfeld (1981) on page 44 about undergraduates at the University of Michigan.

$$y_i = \beta_1 + \beta_2 s_i r_i + \beta_3 (1 - s_i) r_i + \beta_4 s_i d_i + \beta_5 (1 - s_i) d_i + \varepsilon_i \quad i : 1, 2, \cdots, 32$$

where y_i denotes rent paid per person, d_i distance from campus in blocks, r_i number of rooms per person, s_i is a dummy variable representing gender (one for male and zero for female).

The restrictions considered on the β's are

$$\beta_2 \geq 0, \ \beta_3 \geq 0, \ \beta_4 \leq 0, \ \beta_5 \leq 0$$

$$H_{BKS} = \begin{pmatrix} 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad G_{BKS} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

$$H_{G} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad G_{G} = \begin{pmatrix} +\infty \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$H_{S} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad H_{S'} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Hyperparameters:

$$a = b = 0.001, \quad \mu_S = \begin{pmatrix} 130.0 \\ 123.0 \\ 0.0 \\ -1.153 \end{pmatrix}, \quad \mu_{S'} = 37.63,$$

$$\Sigma_S = (X_S^T X_S)^{-1}, \quad \Sigma_{S'} = (X_{S'}^T X_{S'})^{-1}.$$

For Geweke method:

$$\mu_\beta = (37.63, 130.0, 123.0, 0.0, -1.153)^T, \quad \Sigma_\beta = (X^T X)^{-1}$$

Parameters	BKS-method	Geweke's method
β_1	37.7037 (5.1998)	37.8017 (21.9386)
β_2	134.8952 (9.8633)	134.9228 (24.4277)
β_3	122.7444 (9.7056)	122.7571 (25.3568)
β_4	-0.6447 (0.5692)	-0.6573 (0.5787)
β_5	-1.1448 (0.3872)	-1.1522 (0.3938)
σ^2	1316.165 (349.1228)	1323.586 (353.8908)

Table : Estimates and SD by using the BKS-method and Geweke's method based on 10^4 Gibbs paths.

Amin, M., Qasim, M., Amanullah, M., and Afzal, S. *Performance of some ridge estimators for the gamma regression model*, Statistical Papers, 2005.

G. G. Judge, and T. Takayama, Inequality restrictions in regression analysis. Journal of the American Statistical Association, 61 (1966) 166-181.

L. A. Escobar and B. Skarpness, The bias of the least squares estimator over interval constraints, Economics Letters 20 (1986) 331–335.

L. A. Escobar and B. Skarpness, Mean squared error and effciency of the least squares estimator over interval constraints, Communications in Statistics 16 (1987) 397–406.

K. Ohtani, The MSE of the least squares estimator over an interval constraint, Economics Letters 25 (1987) 351–354.

J. Geweke, Exact inference in the inequality constrained normal Linear regression model, Journal of Applied Econometrics,1 (1986) 127–141.

J. Geweke, Bayesian inference for linear models subject to linear inequality constraints. modeling and prediction: Honouring Seymour Geisser, eds. W. O. Johnson, J. C. Lee, and A. Zellner, New York, Springer (1996) 248–263.

R. Ghosal and S. K. Ghosh, Bayesian inference for generalized linear model with linear inequality constraints. Computational Statistics and Data Analysis, 166 (2022) p.107335.

G. Rodriguez-Yam, R. A. Davis and L. L. Scharf, Efficient gibbs sampling of truncated multivariate normal with application to constrained linear regression. In Technical report, Colorado State University. Unpublished manuscript (2004).

Li Y. and Ghosh S.K., *Efficient sampling methods for truncated multivariate normal and student-t distributions subject to linear inequality constraints*. Journal of Statistical Theory and Practice, **9**, 712–732, (2015).

Pindyck R.S. and Rubinfeld D.L., *Econometric models and economic forecasts*, Second edition, McGraw-Hill: New York, (1981).

Reangsephet, O., Lisawadi, S., Ahmed, S.E., *Adaptive estimation strategies in Gamma regression model*, Journal of Statistical Theory and Practice, 14 (1), Article number 8. 2020.

Asar, Y. and Korkmaz, M., *Almost unbiased Liu-type estimators in gamma regression model*. Journal of Computational and Applied Mathematics, 403, 113819, 2022.

Thank You for Your Attention