GeM Sellers' Rating 2021 for Products

Determination of Weights of Variables for Computation of Overall Rating of the Sellers on the Government e-Marketplace (GeM) Using Data-Driven Algorithm

The Government e-Marketplace (GeM)
Ministry of Commerce and Industry
Government of India
2021

Index

Executive Summary

The Government e-Marketplace (GeM) is a path-breaking revolution in India's public procurement. The GeM is the national public procurement portal for an online, end-to-end marketplace for open, efficient, and transparent procurement of goods and services by Central and State Government organizations. To enable buyers to select credible sellers and establish trust in procurement, GeM needs a robust rating system for sellers. It is critical to arriving at a sound and robust methodology to determine the weights for different variables, which will impact sellers' ratings and substantially affect their performance and competitiveness. To the best of our knowledge, no other Government-to-Government e-Marketplace employs such a rating system incorporating transaction-level variables, users', and buyers' feedback data. In this study, we fill the gap of the non-availability of a data-driven approach to derive the weights of variables by combining transaction-level, *users*', and *buyers*' feedback data for Government-to-Government e-Marketplace.

In this study, our objective is to develop a methodology for suggesting weights and determination of weights of the variables in computing the sellers' overall rating. The principal component analysis (PCA), a multivariate technique, has been employed to develop an objective methodology to obtain weights of different variables, namely *delivery*, *reliability*, *quality*, *users*' feedback, and *buyers*' feedback to calculate sellers' rating on GeM. The overall rating is further adjusted by the penalty score derived from the "*incident*" to arrive at the overall final rating. We derived the *delivery*, *reliability*, and *quality* from the transaction-level data. *Users' feedback*, *buyers' feedback*, transaction-level data, and "*incident*" are likely to reveal the sellers' intrinsic rating. The rating system can address the opaqueness and trust-deficit, if any, of the public procurement system to a great extent by deploying the rating model developed in this study.

1. Introduction

Public procurement, *i.e.*, government agencies' purchase of goods and services, is an important economic activity for countries worldwide. 20-22% of India's gross domestic product (GDP) comes from public procurement. Considering the Indian economy's size as USD 2.7 trillion, this amounts to a staggering USD 500 billion annually. Out of this, about USD 100 billion of goods and services are amenable to procurement from a marketplace like Government e-Marketplace ("GeM" or "the platform") annually. We cannot underestimate the importance of driving efficiency, transparency, and effectiveness in this crucial government function in this context. It is a prerequisite to promote resource efficiency and prevent mismanagement and corruption in public procurement processes. With this in mind, India's Government adopted the mandate for a transparent, cashless, contactless, and paperless digital initiative to streamline public procurement on August 9, 2016. The Government of India decided that a platform called the GeM will be developed and implemented as a comprehensive national public procurement portal. The intent was to enable the procurement of goods and services required by Central and State government organizations. To streamline the public procurement of goods and services, India sets up the GeM platform. Over the past four years, GeM has steadily emerged as an exemplar and revolutionized

India's public procurement regime through its online, end-to-end marketplace. GeM facilitates the open, efficient, and transparent procurement of goods and services by central and state government agencies in India. We find that the number of product categories has increased to 6,581 from about 3,486 a year ago.

Similarly, the number of service categories has increased from 75 to 121. The number of buyers has increased from 0.035 million to 0.045 million. We also find that the number of sellers has gone up from 0.229 million to 0.388 million. The number of sellers in the category of micro, small, and medium enterprises (MSME) has increased from 0.038 million to 0.1 million, the number of products has increased from 0.934 million to more than 1.8 million, and the cumulative gross merchandise value has increased from INR 300.31 billion to INR 543.36 billion.

The GeM platform used technology, analytics, and digitization of processes to transform the legacy procurement systems previously in place. It developed a disruptive marketplace model in what is perhaps one of the most complex change management processes in government technology in India and globally. The Government of India's commitment to the three pillars critical to realize the ambitious vision of a genuinely self-sufficient nation or "*Atmanirbhar Bharat*" (self-reliant India) underlined the design and

development, and implementation of the platform. Firstly, it promotes inclusivity by dramatically reducing the cost of doing business and providing typically under-served sellers pan-Indian access to buyers from Government agencies. Secondly, it enhances the procurement process' usability and transparency by relying on cutting-edge technologies and data analytics. Moreover, thirdly, it also aids in increasing efficiencies and reducing costs.

Since its inception, GeM has proven to be transformative in streamlining and improving public procurement processes in the four years. The Gem has enabled increased compliance through its focus on transparent and rule-based processes. Besides, GeM made the enforcement of complex policies like the public procurement (preference to make in India) (PPP-MII) and public procurement policy for micro and small enterprises (PPP-MSE), which had been challenging to implement and monitor in a manual system efficiently. The Government is committed to the procurement process and experience through GeM for sellers and government buyers to pursue its foundational pillars of inclusivity, usability, transparency, and efficiency. In this context, GeM has a rating system for sellers on the platform, which has a dual purpose: one, to enable buyers to select credible sellers, and two, sellers to establish trust in procurement.

The first rating system is available in the GeM framework document. Initially, we determine the rating system for sellers based on four variables: *delivery* time ("*delivery*"), *reliability*, *quality* of order fulfillment ("*quality*"), and *coverage*. We define these variables in Section 2.1 below. During extensive deliberations with the relevant stakeholders within GeM, the variable "*coverage*" allocated a higher rating to more prominent sellers with a pan-Indian presence. The variable, *coverage*, put smaller sellers in the early stages of operation or do not have reach across states, at a disadvantage. The promotion of inclusivity, MSMEs, and start-ups is key to the fundamental commitment of the GeM platform. It is critical to enable start-ups and smaller sellers to participate on the GeM platform and reap its benefits on an equal footing and promote their competitiveness as far as possible. We exclude the "*coverage*" variable from the platform's overall sellers' rating system in this context. Additionally, more recently, the rating system has been modified to include users' feedback ("*Users'* ") and buyers' feedback ("*buyers'* ") as two additional contributing variables - an essential component for robust marketplaces. Therefore, in the computation of overall rating, *users'* feedback and *buyers'* feedback are also included along with three transaction-level variables, i.e., *delivery*, *quality*, and *reliability*.

It is critical to arriving at a sound and robust methodology to determine the weight for each of these five variables, which will impact sellers' ratings and substantially affect their performance and competitiveness. In this study, our objective is to develop a methodology for suggesting weights and determination of weights of the variables mentioned above in computing the sellers' overall rating.

2. Methodology

2.1 Description of variables

In this study, we used transaction-level data obtained from the GeM platform. We derived the data about *delivery*, *reliability*, and *quality* from the transaction-level data of 2017-2020. We collected the users' and buyers' feedback data during the first and second quarters of 2020. The variables under consideration *i.e.*, "*delivery*", "*reliability*", "*quality*", "*users'*" feedback, "*buyers'*" feedback and "*overall rating*" are denoted by *X1*, *X2*, *X3*, *X4*, *X5* and *Y* respectively. Additionally, we adjusted the overall rating based on the reporting of incidents, *i.e.*, the seller's non-compliance with the terms and conditions agreed between the seller, buyer, and the GeM platform. We present a brief description of the variables below:

Delivery: **This variable quantifies the seller's effectiveness in adhering to the service levels of GeM in terms of the delivery schedule committed to them. We calculated rating at each completed transaction level. The higher the value of this variable, the better it is for both the buyer and the seller. We compute the variable entitled "*delivery*," as per Table A1, presented in the Appendix. The computation is to be carried out daily using the entire data for the seller till that date.**

Reliability: **This variable quantifies the ability of a seller to accept orders from buyers. A metric, known as the "*acceptance ratio*", is used to compute the reliability. We present the formula of acceptance ratio in Eq.1.**

$$Acceptance\ Ratio = 1 - \frac{Number\ of\ orders\ not\ honored, i.e., declined\ or\ cancelled}{Total\ number\ of\ orders\ placed.} \quad (1)$$

The higher the acceptance ratio, the higher is the reliability of the respective seller. A higher acceptance score and more excellent reliability are beneficial for both buyers and sellers on the platform. We describe different situations for arriving at the sellers' *reliability* in Table A2 presented in the Appendix.

Quality: **In GeM's rating system, this variable quantifies the quality using a metric called the "*rejection ratio*." When a seller delivers a product to a buyer, the buyer is likely to accept it if it meets the specifications indicated by the buyer**

and the quality expectations for the product. Hence, we compute the rejection ratio using Eq. 2.

$$Rejection\ Ratio = \frac{Quantity\ rejected\ by\ the\ buyer}{Quantity\ ordered} \quad (2)$$

A lower rejection ratio implies better adherence to quality. Therefore, a lower rejection ratio is more advantageous in transactions for both buyers and sellers. We compute the variable entitled "*quality*" using the rejection ratio in Table A3 presented in the Appendix.

Users' feedback: This variable captures a user's overall satisfaction level. Users submit ratings of their experience with a seller through the GeM platform. This rating is on a 6-point Likert-like scale, where 5 indicates the highest level of satisfaction, and 0 indicates the lowest level of satisfaction. It is pertinent to examine how many users' feedback ratings will be statistically adequate for rating a seller reliably in this context. Typically, a higher sample size lends greater precision to any statistical conclusions and ratings. However, considering the platform's nature and the varied nature of sellers' goods and services, it could take several months to obtain a sufficiently large and representative dataset of user ratings. The generation of a sufficient number of users' ratings could potentially function to the detriment of a relatively new

seller and adversely impact their overall rating. Conversely, it is also essential to compute the rating as reliably and accurately as possible, which may not be conceivable if the number of data points is below a certain threshold.

To strike a balance for these two conflicting needs, we propose the computation of the rating based on the statistical concept of the "sampling distribution of the sample mean." The sampling distribution of the sample mean is normally distributed with mean as the population mean and variance as the population variance divided by the number of observations. Hence, the sample mean standard error is the population standard deviation divided by the square root of the sample size. Therefore, the higher the sample size, the lower is the standard error. A lower standard error implies a better precision of the estimator. This concept is used widely in statistical computations, particularly in the design of experiments, among other areas, where the number of replications is required to be determined optimally. In the experimental design, a replication of three is considered adequate in many experiments, particularly in a field trial, where it is challenging to arrange many subjects for an investigation. With this fundamental concept in view, we have considered a minimum number of required sample size as three. We believe a minimum of three sets of feedback from different users is adequate for the computation of users' ratings. We compute the simple average and median of three or more

ratings. We retain the higher of these two values as the user rating. The higher the user rating, the better it is for the buyer and the seller.

Buyers' feedback: In the GeM platform, buyers are different from the users. Buyers are the administrative department of concerned Ministries and Organizations. Buyers' rating is on a 6-point Likert-like scale, where 5 indicates the highest level of satisfaction, and 0 indicates the lowest level of satisfaction. The methodology of collecting this data is the same as mentioned under the users' feedback.

2.2. Incident history and its impact

The failure to honor the terms and conditions of the platform is known as an "*incident*." The reporting of "*incidents*" concerning sellers can also affect their overall ratings. When contracting through the platform, buyers and sellers must comply with the general terms and conditions of GeM and those specific to their agreement. Non-compliance leads to a negative impact on the vendor's overall rating. This non-compliance can also lead to the suspension of sellers. The suspension can be three types: serious, severe, and grave incidents (Please refer to GeM Incident Management Policy). Each suspension will contribute to an overall negative impact as defined in the matrix below (Table 1). The maximum negative impact of this variable is capped to 2.5 rating points.

Table 1 Impact of incidents

Sl. No.	*Conditions*	*Negative Impact on Overall Rating*
1	Each Suspension for Grave	1 Rating Point
2	Each Suspension for Severe	1 Rating Point
3	Each Suspension for Serious	1/3 (0.33) Rating Point

Auto incidents will be generated against the seller by the GeM platform in case of deviations as per incident management policy.

2.3 Proposed model for calculating the overall rating

In this study, we attempt to compute the weights assigned to the five variables to calculate the *overall rating*. The proposed model is:

$$\boldsymbol{Overall\ rating\ (Y) = w1X1 + w2X2 + w3X3 + w4X4 + w5X5} \quad \textbf{(3)}$$

The weights w1, w2, w3, w4, and w5 are expressed in the fractions and are required to be determined. We compute the *overall final rating* by adjusting the impact of the incident on the *overall rating*.

2.4. Pre-processing of data and the theoretical background

In the pre-processing stage, missing values were identified and appropriately coded to carry out the computation without the missing values. We employ the principal component analysis (PCA) to compute the weights. We present a brief description of the PCA below:

Principal component analysis: **Principal component analysis (PCA) is a multivariate technique in which we simultaneously analyze all the variables. In the PCA, we compute a set of orthogonal eigenvectors of the variables' correlation matrix (say R). The matrix of principal components (say Z) is the product of the matrix of the variables under consideration with the eigenvector matrix. The first principal component (Z1 or PC1) accounts for the most significant percentage of the data set's total variability. The second principal component (Z2 or PC2) accounts for the second-largest percentage of the total variability present in the data set, and so on. We present the methodology here:**

We obtain the eigenvectors as the columns of the orthogonal matrix in the spectral decomposition of the correlation matrix, R. As R is a symmetric matrix, there exists an orthogonal matrix V such that $V^TRV = L$ or, equivalently, $R = VLV^T$, where L is a diagonal matrix whose diagonal elements are the eigenvalues, and T denotes the transpose of a matrix. We can consider the correlation matrix as the three matrices' product - the matrices of eigenvalues and corresponding eigenvectors. The eigenvectors are the columns of the matrix V. Using the linear combinations of the original variables; we calculate the principal component scores using the following equation:

$$Z = XV \qquad (4)$$

Here, Z is the matrix of principal components scores (n x k), X is the standardized data matrix (n x p), and V is the matrix of eigenvectors (p x k). The principal components are uncorrelated, with decreasing variance. The proportion of the total variability due to the i^{th} principal component is given by:

$$\textit{The proportion of total variance due to ith component} = \frac{\lambda_i}{\lambda_1 + \lambda_2 + \cdots + \lambda_k} \quad (5)$$

where, $\lambda_1, \lambda_2, \ldots, \lambda_i, \ldots, \lambda_k$ are the eigen values corresponding to the ith principal component. The principal component analysis aims to explain the maximum amount of variance with the least number of components. We employ the principal component analysis using the following steps:

Step 1 (Normalization of variables): Let, x_{ij} denotes the observation of the ith seller (vendor) of the j^{th} variable. Let us carry out a normalization transformation using the minimum-maximum transformation, and Nx_{ij} denotes the transformed values.

$$Nx_{ij} = \frac{x_{ij} - Min(x_{ij})}{Max(x_{ij}) - Min(x_{ij})} \quad (6)$$

Step 2 (Computation using the PCA): Having transformed (normalized) the four variables using the above normalization technique, we run PCA on the

transformed data set using the following command sequences in Minitab 19 statistical software:

Stat>Multivariate>Principal components. (7)

3. Results

We presented the results of the PCA in Table A4. The results depicted in Table A4 showed that the five principal components (PCs) thoroughly explained the total variation present in the data. The first PC explains 33.4% of variability; the second PC explains 27.4%; the third PC explains 18.6%; the fourth PC explains 13.1% of the variability, and the fifth PC explains 7.5% of the variability. The results presented in Table A4 depict that the loading on *delivery* is very high for PC1, PC2 is on *the buyer*, PC3 is on *users'*, PC4 is on *quality, and PC5 on reliability*.

We assign the weights of 0.33, 0.27, 0.19, 0.13, and 0.08 to the *delivery*, *buyers'*, *users'*, *quality*, and *reliability*. The determined weights (approximated to the whole numbers) obtained from the PCA are 33%, 27%, 19%, 13%, 8% for *delivery*, *buyers' feedback*, *users' feedback*, *quality*, and *reliability*. We present the weights of variables for computing the overall rating in Table 2.

Table 2 Weights of variables

	Transaction level data			Feedback data	
	Delivery	**Reliability**	**Quality**	**Users'**	**Buyers'**
Model 1	**0.33**	**0.08**	**0.13**	**0.19**	**0.27**
Model 2	**0.41**	**0.21**	**0.12**	**0.26**	
Model 3	**0.40**	**0.34**	**0.17**		**0.09**
Model 4	**0.54**	**0.30**	**0.16**		

We can select a particular model for computation of the *overall rating* depending on the availability of data. For example, we can employ Model 1 if data on all the variables are available. We present here an example of the computation of the *overall rating*. As an example, let the hypothetical individual rating scores of all the variables of a seller are as given in Table 3 below:

Table 3 Hypothetical individual rating scores of a seller

Variable	Rating	Variable	Rating
Delivery	**2.23**	**Buyers'**	**4.0**
Reliability	**4.0**	**Users'**	**5.0**
Quality	**4.34**		

Here, we have data on all five variables. Hence, let us apply Model 1.

$$Overall\ Rating$$

$$= 0.33 * 2.23 + 0.08 * 4.0 + 0.13 * 4.34 + 0.19 * 5.0 + 0.27 * 4$$

$$= 0.7359 + 0.32 + 0.5642 + 0.95 + 1.08 = 3.6501$$

Hence, the overall rating of the seller is 3.65 out of 5. The overall rating is to be rounded to two decimal places.

We apply any history of an "*incident*" about the seller on 3.6501 before arriving at the *overall final rating*. If there is no history of an "*incident,*" then the seller's *overall final rating* is 3.65 out of 5. We'll carry out the computation based on all available data daily, and the "*incident*" once occurred shall continue for one year from the date of occurrence in the GeM platform. If for this seller, the incident score is 2.33, the overall final rating will be (3.65-2.33) = 1.32 out of 5.

In situations other than Model 1 to Model 4, *individual variable rating (partial rating)* can be reported and displayed on the system. However, the *overall rating* is to be reported as NA (Not Available). If such sellers are penalized as a result of incidents raised against them by the platform or otherwise, then the cumulative penalty points against that seller will also be reported and displayed under penalty due to incidents as per Table 1.

4. Conclusions

The sellers' computation of rating was carried out using transaction-level variables (delivery, reliability, and quality), feedback variables (users' and buyers'), and penalty due to incidents. The weights for transaction-level variables and feedback variables are obtained using principal component analysis. The computation of rating is to be carried out daily using the completed order.

Appendix

Table A1 Delivery

Sl. No.	Conditions	Rating
1	Delivered on or before the scheduled time	5
2	Delivered after the scheduled period and within ten days' delay	4
3	Delivered after the scheduled period and within 11-20 days' delay	3
4	Delivered after the scheduled period and within 21-30 days’ delay	2
5	Delivered after the scheduled period and within 31-45 days’ delay	1
6	Delivered after 45 days	0

Table A2 Reliability

Sl. No.	Conditions	Rating
1	Acceptance ratio is 100%	5
2	Acceptance ratio is more than equal to 90% and less than 100%	4
3	Acceptance ratio is more than equal to 75% and less than 90%	3
4	Acceptance ratio is more than equal to 50% and less than 75%	2
5	Acceptance ratio is more than equal to 25% and less than 50%	1
6	Acceptance ratio is below 25%	0

Table A3 Quality

Sl. No.	Description	Rating
1	Rejection ratio is 0%	5
2	Rejection ratio is more than 0% to less than equal to 5%	4
3	Rejection ratio is more than 5% to less than equal to 10%	3
4	Rejection ratio is more than 10% to less than equal to 20%	2
5	Rejection ratio is more than 20% to less than equal to 40%	1
6	Rejection ratio is more than 40%	0

Table A4 Results of PCA (Eigen analysis of the correlation matrix)

Eigen value	1.67	1.37	0.93	0.66	0.37
Proportion	0.334	0.274	0.186	0.131	0.075
Cumulative	0.334	0.608	0.794	0.925	1.000
Variable	***PC1***	***PC2***	***PC3***	***PC4***	***PC5***
Delivery	0.67	0.021	0.213	0.263	-0.661
Buyers' feedback	0.187	0.704	0.218	0.445	0.469
Users' feedback	0.278	0.171	-0.943	0.06	0.007
Quality	0.614	-0.053	0.131	-0.668	0.397
Reliability	-0.244	0.687	0.016	-0.532	-0.431

Glossary

Individual Variable Rating: It is defined as the individual variable's rating score based on its description. An individual variable rating can be obtained for delivery, reliability, quality, users' and buyers' feedback rating. In case the overall final rating is not available, an individual variable rating can be reported and displayed on the system. Individual Variable Rating is also known as *Partial Rating*.

Overall Rating: It is defined as the rating score obtained by the developed rating model, a linear combination of rating variables multiplied by its weight and summed over all the variables.

Overall Final Rating: It is defined as the overall rating score adjusted by the penalty score derived from the incident(s). In the absence of any incident, the *overall rating* and the *overall final rating* are the same.