Mehrjoo, Saeed; Bashiri, Mahdi

Article
An application of principal component analysis and logistic regression to facilitate production scheduling decision support system: An automotive industry case

Journal of Industrial Engineering International

Provided in Cooperation with:
Islamic Azad University (IAU), Tehran

Suggested Citation: Mehrjoo, Saeed; Bashiri, Mahdi (2013) : An application of principal component analysis and logistic regression to facilitate production scheduling decision support system: An automotive industry case, Journal of Industrial Engineering International, ISSN 2251-712X, Springer, Heidelberg, Vol. 9, pp. 1-12, http://dx.doi.org/10.1186/2251-712X-9-14

This Version is available at:
http://hdl.handle.net/10419/147155

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes.
You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

http://creativecommons.org/licenses/by/2.0/
An application of principal component analysis and logistic regression to facilitate production scheduling decision support system: an automotive industry case

Saeed Mehrjoo and Mahdi Bashiri

Abstract
Production planning and control (PPC) systems have to deal with rising complexity and dynamics. The complexity of planning tasks is due to some existing multiple variables and dynamic factors derived from uncertainties surrounding the PPC. Although literatures on exact scheduling algorithms, simulation approaches, and heuristic methods are extensive in production planning, they seem to be inefficient because of daily fluctuations in real factories. Decision support systems can provide productive tools for production planners to offer a feasible and prompt decision in effective and robust production planning. In this paper, we propose a robust decision support tool for detailed production planning based on statistical multivariate method including principal component analysis and logistic regression. The proposed approach has been used in a real case in Iranian automotive industry. In the presence of existing multisource uncertainties, the results of applying the proposed method in the selected case show that the accuracy of daily production planning increases in comparison with the existing method.

Keywords: Principal component analysis, Logistic regression, Production planning control, Decision support system

Introduction
Effective planning and control of production processes are usually seen as key to the success of a manufacturing company. During the last 50 years, both academic institutes/universities and industries have put great effort into developing and designing successful approaches and methods for manufacturing planning and control. Indeed, the methods and approaches of how to plan and control production have been changed over time. This occurs in line with changes in customer requirements and technology improvements (Vollmann et al. 2005).

Detailed production scheduling is an extremely complex problem (Brucker 2007) wherein most cases are considered NP-hard (Günther and van Beek 2003). In order to deal with complexities and uncertainties, a detailed production scheduling system should be equipped with all the necessary decision support tools for rendering production problems visible within a planning period and shift dispatching control from the foremen to the planner (Sotiris et al. 2008). According to Simchi-Levi et al. (2008), the decision support system (DSS) is an analytical tool to aid operations and production planning. The DSS can range from simple tools to expert systems. The DSS helps to solve the problems such as network planning to tactical planning all the way to daily operational problems. Thus, the effective DSS can help managers or production planners to manage uncertainties and achieve better results in daily fluctuations.

The stimulus for this work has been to understand whether or not the historical daily shop floor data can be used for creating more robust daily production plan. Moreover, the paper studies the feasibility of using the multivariate statistical analysis of daily shop floor data as an appropriate solver tool for detailed production scheduling decision support system. In order to answer these, we represent an Iranian automotive case of detailed production planning in applied material requirement planning (MRP) system. The results may not be generalized to JIT and lean manufacturing principles which have a
pull approach of planning and control of production. The rest of the paper has been organized as follows: in next section, the related literature has been reviewed, whilst the problem has been defined and the selected case has been presented in the ‘Case problem statement’ section. The ‘Methodology of problem analysis’ section has outlined the proposed multivariate DSS as subsequent specifications arising from the ‘Case problem statement’ section. The ‘Proposed multivariate DSS method’ section has demonstrated the implementation results, and finally, in the ‘Conclusions’ section, the conclusions have been drawn and further research efforts have been mapped out.

Literature review
The production planning control models can be characterized in a variety of approaches, but most common categorizations are specific to the application areas (Brucker 2007). In this article, we have classified the production planning problem solving approaches in two groups in terms of their environment and condition: unconditional/deterministic analytic production planning and production planning under uncertainties (see Table 1).

In the case of unconditional analytic models, we are referring to models which are simplifications of a real system in terms of mathematical expressions and can be solved by exact or heuristic methods. The literature on the exact algorithms in scheduling and production planning problems is extensive. A thorough review of scheduling problems, modeling approaches, and solution methods can be found in Brucker (2007), Framinan and Ruiz (2010), and Ribas et al. (2010). Whereas fluctuations and uncertainties have key roles in real world, deterministic, or exact approaches, such that branch and bound (Yao et al. 2012) and mixed integer linear programming (Maravelias and Sung 2009) are seldom applicable in actual shop floors since they may only solve small-scale problems with distinct parameters and scale of time.

Several heuristics and hybrid methods are recommended in the literature (Jourdan et al. 2009). Ribas et al. (2010) have classified the approximate methods into constructive and improvement heuristics. However, the application of heuristic algorithms is believed to be more applicable for real shop floors due to their lower computations, but they are still limited by the dimension of problems and uncertainties. Therefore, their implementation should be coupled with some decision support tools to aid the production planners (Ross and Bernardo 2011; Sotiris et al. 2008).

To cope with uncertainties in production control, it is worth to investigate a new customized framework for planning and scheduling under uncertainty. Another challenging issue is to investigate the ways of controlling a large number of uncertain parameters. Hence, scheduling under uncertainty has received a lot of attention in recent years (e.g., Hatzikonstantinou et al. 2012; Vargas and Metters 2011; Torabi et al. 2010; Verderame and Floudas 2009). Uncertainty can be derived from many aspects, such as demand or product orders, alternation or priority of orders, equipment failures, resource changes, and processing time variability. To adapt uncertainties during the manufacturing process, the proposed methods are divided into two main groups: reactive scheduling and preventive scheduling (Ayug et al. 2005). Simulation approach is able to analyze the behavior of the environment when it is characterized by several constraints and uncertainties (Rolo and Martinez 2012; Volling and Spengler 2011; Jahangirian et al. 2010). In these approaches, the outcomes of the simulation software can be used in preventive scheduling and decision support systems, but they need a great deal of efforts to make a practical schedule by some expert production planners.

By the emersion of enterprise resource planning (ERP) systems, the utilization of data becomes more important in production planning and control (PPC). The incredible wealth of available data in SCM and PPC software raises the question of how to help decision makers in harnessing the organization. The answer to this question has defined the production activity control (PAC) subsystem at the lowest level of MRPII (Vollmann et al. 2005). By means of the PAC system, the sequence of the orders is defined with their release and due times. In fact, PAC cannot take into account the real state of the production environment, and it may produce unrealistic or impractical production plan.

Whereas the MRP-based system cannot follow the large number of shop floor fluctuations, production managers bow to the inevitable complex task of scheduling/rescheduling at the shop floor control. Poor production control may cause serious problems to a firm’s ability to meet production requirements and constraints. Many researches have focused on developing DSS tools to face this problem (e.g., Ko and Wang 2010; Caricato and Grieco 2009; Mok 2009; Farrella and Maness 2005; McKay and Wiers 2003). These tools are concerned as complementary applications to the ERP/MRP software.

Unfortunately, few success stories have been reported on creating production planning and logistics in a real factory, and there are still many challenges that remain (McKay and Black 2007). In the absence of one sole issue for PPS success or failure (McKay and Wiers 2003, 2004), one potential issue related to the failure of a planning system is the lack of information system and DSS tools for detailed production planning. This was the first insight obtained from this case study (see Table 1).

Meanwhile, combination use of statistical analysis with other methods to control the uncertainties in real condition decision making has been proposed by some
Table 1 Classification of related literatures

| Approach                      | Deterministic production planning | Production planning under uncertainties |
|-------------------------------|-----------------------------------|------------------------------------------|
|                               | Exact methods                     | Simulation                               |
| Literature                    | Brucker (2007)                    | McKay and Wiers (2003)                   |
|                               | Maravelias and Sung (2009)        | Sotiris et al. (2008)                    |
|                               | Framinan and Ruiz (2010)          | Volling and Spengler (2011)              |
|                               | Ribas et al. (2010)               | Mok (2009)                               |
|                               | Yao et al. (2012)                 | Rolo and Martinez (2012)                 |
|                               |                                  | Caricato and Grieco (2009)               |
|                               |                                  | Ko and Wang (2010)                       |
|                               |                                  | <-------------------------------------->
|                               | Heuristic methods                | Statistical analysis                     |
|                               | Sotiris et al. (2009)            | Hatzikonstantinou et al. (2012)          |
|                               | Jourdan et al. (2009)            | Aytug et al. (2005)                      |
|                               | Jangangirian et al. (2010)       | Farrella and Maness (2005)               |
|                               | Ribas et al. (2010)              |                                        |
|                               | Yao et al. (2012)                |                                        |
|                               |                                  | <-------------------------------------->
|                               |                                  | Hybrid/combinational methods             |
|                               |                                  | Cunha and Wiendahl (2005)               |
|                               |                                  |                                        |
|                               |                                  |                                        |

**Comments**
- Seldom applicable in actual shop floors since they may only solve small-scale problems with distinct parameters and scale of time.
- Simulation needs a great deal of efforts to make a practical schedule by some expert and expensive production planners.
- DSS tools are concerned with complementary applications to ERP/MRP software. They are practical with lack of ERP system.
- Statistical and hybrid methods are useful to control uncertainties in real condition and improve the effectiveness of both evaluation and decision making; however, they are not independent and complete tools. They have to be designed for each case problem.

DSS, decision support system; MRP, material requirement planning; ERP, enterprise resource planning.
The main problem of the selected case has been inferred from logistics staff answers to a set of questions and interviews. It is reported that the accuracy of daily production plans is directly affected by some alternate constraints and probable parameters. Hence, either the rescheduling or planning diversity and related extra material handling or extra/shortage parts and production line stop are enviable tasks every day. By their complaint about the alternate decisions to manage stochastic or abnormal events, we have made inferences about the lack of DSS tools to provide a practical detailed production planning in an uncertain condition.

Case problem statement
The case study of detailed production planning has been done in an Iranian automotive manufacturing company. SAIPA Corporation (Tehran, Iran) is a holding company that assembles several types of passenger cars, vans, minibuses, buses, and trucks. As with any other car manufacturing company, the production process followed has a high degree of complexity, coupling the complex bill of materials (BOMs) with equally complex routings that transcend the shop floor boundaries.

The main problem of the selected case has been inferred from logistics staff answers to a set of questions and interviews. It is reported that the accuracy of daily production plans is directly affected by some alternate constraints and probable parameters. Hence, either the rescheduling or planning diversity and related extra material handling or extra/shortage parts and production line stop are enviable tasks every day. By their complaint about the alternate decisions to manage stochastic or abnormal events, we have made inferences about the lack of DSS tools to provide a practical detailed production planning.

Methodology of problem analysis
The following three aspects of the problem have been specified in the analysis of the current situation:

- **Layout and physical constraint.** It focuses on the production flow and is concerned with constraints of layout and any physical limitation in the production lines.
- **Production planning and control system.** It is concerned with the daily activities of the production planners during their detailed production planning in the shop floor control process. Shop floor data in a multi period range was gathered from this aspect of analysis.
- **Uncertainties and stochastic factors.** It is concerned with the source of uncertainties and stochastic factors.

To investigate the mentioned aspects, a mixture of interviews and observation has been applied. The major part of observation and a small part of the meetings were concerned with information about the production process and the production planning control.

In the following two subsections, the basic results concerning the first two aspects of the case problem analysis have been presented. These results are normally used to design the system architecture and functionality as well as the shop floor model of the plant. The results of the third aspect, namely uncertainties, are used to construct a multivariate analysis tool of simplified real production.

Layout and physical constraints
A trim shop is located at the end of the production process. Therefore, it has the highest level of complexity in comparison with subsequent production activity control processes. The main assembly production line is equipped with a conveyor. According to production rate and types of products (seven types), the length of production line is not quite enough for assigning individual locations to keep the minimum stock level of all parts according to the type of product BOM. The logistic area is not available near the line far distance from the main warehouses; thus, the order of completion lead time is long and is influenced by probable accidents.

There is a painted body (PB) stock at the end of the paint shop process. The stock of PB is the same as a single line queue before the entrance of the trim shop, and each PB can be transferred to the trim shop by the sequence of its location. Incapability of selecting the desired PB from the PB stock constrains the production planner to make a daily plan according to the PB color and type sequence. Although the elimination of the layout and physical constraints have been investigated in recent years, due to outstanding required cost and time, the progress of development is not noticeable.

Production planning and control system
Although KANBAN cards and pull production control system have been tried to be applied by production planning and the logistic department, the production control system is still MRP-based. A hierarchical two-level planning framework is used prior to the detailed production scheduling. At the top level, aggregate production planning which controls demand management with a yearly time horizon, has been located. The second planning level which is called midterm planning incorporates a hybrid MRP-PBC approach.

Master production schedule (MPS) outcomes are used to calculate components and material requirements. The final plan is made by revision on a weekly basis using the feedback from the detailed scheduling module. The
weekly plan is released by PAC, and detailed production planning is issued as the daily schedule. The daily schedule is derived from a complicated decision-making process which uses shop floor data, inventory status data, BOM, and sequence of PB in stock. Figure 1 demonstrates this process.

PPC suffers from several sources of inconsistencies as a consequence of incomplete ERP implementation. As a result of its complex production process and lack of information technology (IT) infrastructure, the presented case study during the last few years faced numerous problems concerning violated due dates, accumulated late orders, supernumerary production orders, excessive component inventory, poor releasing policies, and low shop floor visibility. The lack of online and integrated information may cause a misunderstanding of the real condition; thus, the production planner faces some unknown parameters in daily scheduling. In this situation, it is not weird if the daily schedule encounters some mistakes. Although the design and implementation of the ERP software is in progress, production planners cannot wait and do not get along with increasing complexity. They really need some practical tools to help them in perfect decision making.

Uncertainties and stochastic factors
Since there are some line-side space constraints, mixed production suffers from lots of problems and obstacles and forces managers to act on the basis of batch production. Meanwhile, there are many sources of stochastic events and uncertainties that batch production such as demand, process, and supply uncertainties (Peidroa et al. 2009). One of the main sources of stochastic factors that have been identified in this study is derived from the paint shop process and PB stock constraint. Due to some small defects on bodies, some of the PBs are selected to go into a touch-up area, and after doing all necessary reworks, they are transferred to the PB stock line. Almost all of the procedures in the defect inspection process are performed manually through human vision and influenced by stochastic factors. On the other hand, the required rework process times depend on the type and the level of defects which are not really exact and deterministic. Hence, the sequence of painted bodies in the queue of stock line cannot be absolutely defined. Meanwhile, supply uncertainties have a key role in unreliability of the production schedule. Each type of products has special parts which are from different suppliers. The availability of all special parts related to the desired type of products is the other vital information for the production planner to make the daily production schedule. According to our observation, the stock levels of these items are not expected to follow exact patterns.

Proposed multivariate DSS method
The complexity of production planning and control process, stochastic factors, physical constraint, uncertainties, and the shortcomings of the underlying IT
infrastructure would pose significant drawbacks to the current detailed production scheduling. In this light, to the aforementioned production planning process and fully interoperable, both with the PPC system and existing software package, the proposed approach has been developed on the basis of a custom-built DSS using statistical multivariate techniques.

The integrated approach that will be presented introduces facilities to analyze data which are directly unavailable from the current planning system. This approach is introduced through the use of PCA to decrease the dimension of input-independent variables (Aguilera et al. 2006) and the use of logistic regression (LR) to predict the first priority of available and suitable type of product which can be selected to make a practical and effective detailed production schedule. These are used at different steps as shown in Figure 2.

**Shop floor and production plan historical data acquisition**
The manner and logical behavior of the production planner to create a weekly plan or change daily detailed scheduling is an important factor through the practical decision-making process which can be used for finding an effective DSS tool. As answer to the main question of this research, the objective has been to find the statistical analysis appropriate for reducing this logical behavior. Hence, it has been required to collect daily shop floor data and historical data of the daily schedule issued by the production planner. The historical data of PB stock, existing PB quantity, and PB types in paint shop, sale online requests, inventory data, MRP weekly plan, released daily production schedule, and related orders with actual production were the main fields of data that have been collected for this analysis.

**Reduction of inventory data by PCA**
In this study, the collected inventory data sets (warehouse and line side separately) have at least 40 fields related to each types of products. This high volume and dimension of data matrix increase the complexity of analysis. If a substantial amount of the total variance in these data is accounted for by a few (preferably far fewer) principal components or new variables, then these few principal components can be used for interpretational purposes or in further analysis of the data instead of the original variables. PCA can be viewed as a dimensional reduction technique (Sharma 1996), and it is the appropriate technique for achieving the mentioned objective.

The core idea of PCA is to reduce the dimensionality of a data set comprising a large number of interrelated variables while retaining as much as possible the data set variance (Jolliffe 2002). This is obtained by transforming original variables to a new set of variables or principal components that account for much of the variance in the original variables. This process involves the following steps:

1. **Collect Daily Shop Floor & Actual Historical Data:**
   - PB sequence
   - Sales/ Demand
   - Line Side Inventory Data
   - Warehouse Inventory Data
   - MRP weekly plan
   - Actual Production

2. **Collect Daily Production Plans Historical Data:**
   - Released production plan
   - Released Orders
   - Rolling Plans
   - Production Plans Changes times
   - Daily Plans Performance Index

3. **Current Shop Floor Data:**
   - PB sequence
   - Sales/ Demand
   - MRP weekly plan

4. **Current Inventory Data:**
   - Line Side Inventory Data
   - Warehouse Inventory Data

5. **Using PCA on Inventory Data: Data Dimension Decreasing**
6. **Match Shop Floor and Plans Historical Data**
7. **Fit Logistic Regression Model**
8. **Remove correlation or transform data as needed**
9. **Calculate PC by using current Inventory Data**
10. **Practical LR Model**
11. **Validate Model**
12. **Review Model for Other Insights and future conditions**
13. **DSS**

**Figure 2 Process of establishing a multivariate statistical tool for DSS development.**
components ($\xi_i$) which are a linear combination of original ($p$) variables. Due to their properties, they are uncorrelated and are ordered such that the first ($m \leq p$) that are retained contain most of the variation presented in the original data:

$$PC = \{\xi_1, \ldots, \xi_m\}; \quad \xi_i = w_{i1}x_1 + w_{i2}x_2 + \ldots + w_{ip}x_p,$$

where principal component (PC) = $\{\xi_1, \ldots, \xi_m\}$ are the $m$ principal components and $w_{ij}$ is the weight of the $j$th variable for the $i$th principal component.

The reduction in complexity is achieved by performing PCA on collected inventory data. Thus, the original data of inventory can be substituted by PCs, and the new matching table of the shop floor data and corresponding production schedule is established as a contingency table.

### Logistic regression model fitting, validation, and review to improvement

The fundamental question in this research motivated us to understand the logical behavior of the production planner in the decision-making process through daily production scheduling. As illustrated in Figure 2, the historical input/output of the decision process is analyzed and the relationship among them is discovered by logistic regression. In the remainder of this section, we briefly discuss about the basic concept and details of developing the logistic regression model and, finally, the validation procedure and review method for the improvement of this model.

### Definition of variables

To simplify the discussion and interpretation of estimation model, the notation is introduced and variables are defined which can be recognized from collected data. Table 2 shows a code sheet for definition of preliminary selected variables from collected data.

| Variable | Description | Code/values |
|----------|-------------|-------------|
| TYP      | Compressed natural gas equipped CNG | CNG |
|          | CNG and hydraulic steering wheel CNG-H | CNG-H |
|          | Hydraulic steering wheel HYD | HYD |
|          | SABA simple injection system GLXi | GLXi |
|          | Reinforced new body X132 | X132 |
|          | Morvarid hatchback DM | DM |
|          | Other new models NEW | NEW |

**CLR** Color production plan
- W = white
- G = gray
- B = black
- R = red
- S = silver

**PBS** Painted body stock 0-400

**SEQ** Compatibility of PB sequence
- 1 = Low compatibility
- 2 = Fair compatibility
- 3 = Good compatibility
- 4 = High compatibility

**INV1** Warehouse and line-side inventory PC 1
- 0 to 2,000

**INV2** Warehouse and line-side inventory PC 2
- 0 to 1,000

**SRT** Scheduled receipt
- 0 to 1,000

**ACP** Available colored parts
- 0 to 1,000

**RWP** Remainder quantity of weekly plan
- 0 to 5,000

**ESD** Emergency sales/demand
- 1 = Critical
- 2 = Urgent
- 3 = Normal
- 4 = Non-emergency

**ADP** Actual daily product
- 0 to 1,100

**DPP** Daily production plan
- 0 to 1,100

**DPC** Daily production plan capability
- 0 = No
- 1 = Yes

### Basic theory on logistic regression

There are two models of logistic regression to contain binomial/binary logistic regression and multinomial logistic regression. Binary logistic regression is typically utilized when the dependent variable is dichotomous and the independent variables are either categorical or continuous variables (Sharma 1996). Logistic regression is the best to use in this condition. The result of this type of regression can be expressed by a logit function as follows:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \ldots + \beta_k \cdot x_k = \beta_0 + XB,$$

where $\left(\frac{p}{1-p}\right)$ is the odds.

The model can either be interpreted using the logit scale, or the log of odds (the relative probability) can be converted back to the probability such that

$$p = \frac{\exp(\beta_0 + XB)}{1 + \exp(\beta_0 + XB)}$$

In order to calculate the parameters $\beta_0$, $\beta_1$, $\beta_2$, ..., $\beta_k$, the logistic regression transforms the dependent into a logit variable and then uses maximum likelihood estimation. In this paper, logistic regression is used to estimate the daily production planning capability (DPC) from the shop floor data. According to the variables summarized
in Table 2, the logit can be defined for this case as follows:

\[
\text{logit} = \beta_0 + \beta_1 \times \text{TYP}\{1\} + \beta_2 \times \text{TYP}\{2\} + \beta_3 \times \text{TYP}\{3\} + \beta_4 \times \text{TYP}\{4\} + \beta_5 \times \text{TYP}\{5\} + \beta_6 \times \text{TYP}\{6\} + \beta_7 \times \text{INV1} + \beta_8 \times \text{INV2} + \beta_9 \times \text{CLR}\{1\} + \beta_{10} \times \text{CLR}\{2\} + \beta_{11} \times \text{CLR}\{3\} + \beta_{12} \times \text{CLR}\{4\} + \beta_{13} \times \text{CLR}\{5\} + \beta_{14} \times \text{DPP} + \beta_{15} \times \text{PBS} + \beta_{16} \times \text{SEQ}\{1\} + \beta_{17} \times \text{SEQ}\{2\} + \beta_{18} \times \text{SEQ}\{3\} - \beta_{19} \times \text{SRT} + \beta_{20} \times \text{ACP} - \beta_{21} \times \text{RWP} - \beta_{22} \times \text{ESD}\{1\} + \beta_{23} \times \text{ESD}\{2\} - \beta_{24} \times \text{ESD}\{3\}.
\]

To find out how effective the model expressed in Equation 3 is, the statistical significance of individual regression coefficients is tested using the Wald chi-square statistic. Goodness-of-fit test assesses the fitness of a logistic model against actual outcomes. Hosmer-Lemeshow test is an inferential goodness-of-fit test which is utilized in this paper. Meanwhile, the consequent predicted probabilities can be revaluated with the actual outcome to determine if high probabilities are indeed associated with events and low probabilities with non-events. The readers are referred to Bewick et al. (2005) and Hosmer and Lemeshow (2000) for more information about the assessment of fitted model.

Predicting capability of daily production planning

The fitted model, which has successfully passed the goodness-of-fit tests, can be used to calculate the predicted Logit (probability) of DPC for a given value of shop floor data. For example, assume that at the end of the working day, the production planner wants to make a decision about tomorrow's production plan and would like to predict the capability of a given production schedule. At first, using the PCA method ('Reduction of inventory data by PCA' section), the inventory level of line-side and related warehouses can be estimated by two principal components (INV1 and INV2), and then, according to the shop floor data, the amount of other independent variables (TYP, PBS, RWP, ESD, and DPP) are defined. Therefore, the probability of response variable (DPC) can be calculated by Equation 4:

\[
p = \frac{\exp\left(\text{logit}\right)}{1 + \exp\left(\text{logit}\right)}.
\]

Customized DSS to facilitate detailed production scheduling

Predicting the capability of daily production planning facilitates the decision making of the PPC system, and as a result, the customized DSS can be defined and applied. The new hierarchical planning framework is depicted in Figure 3. The main procedural sequence does not exhibit any remarkable change in comparison with the current PPC process (Figure 1). The MPS calculates long-term end item needs and feeds the PPC system which creates the production order backlog according to MRP procedures. The weekly production plans are issued by the MRP module and feeds detailed scheduling.

The proposed multivariate method contributes in the DSS module which is denoted in Figure 3 in the dashed box. As we have described in the previous sections, the online shop floor data is used by this customized DSS tool and the predicted amount of DPC index is calculated. This production planning capability index can facilitate the decision-making process of detailed scheduling. The production planner can typically run this customized DSS at the beginning of each planning period (commonly one working day), and after making the decision about final changes on the detailed schedule, data of production order is extracted. When detailed scheduling is finalized, the production orders are handed down to the foremen for beginning of production. If dynamic events take place (e.g., a machine breaks down, a rush order arrives, or a subcontractor violates due dates), the planner reschedules to accommodate them.

Numerical experiment and results

According to the defined variables, the 42-week shop floor and inventory data have been collected. Every day, the line-side inventory level of special and important parts as well as warehouse inventory level have been recorded. Table 3 illustrates the sample data which were recorded on the first and second days in the warehouse. Data were collected over a period of 8 months and included PB stock, sale online requests, inventory data, MRP weekly plan, inventory status data schedule, and actual production.

According to the data reduction method described in the ‘Reduction of inventory data by PCA’ section, the following PC scores can be derived by applying PCA. In this study, the result of PCA shows that the first two principal
component variables account for about 90% of the total variance of data, and the screen plot shows that the appropriate number of PCs is two. Using these PCA scores, the PC formulas are defined, and then the amount of PCs by the exact amount of each variable every morning can be calculated. By the results of data reduction, we can make a new shop floor control (SFC) data which can be used for calculating logit more easily and practically. Table 4 shows the sample format of SFC data table which must be created for logistic regression analysis. Using Equation 4,

Table 3 Sample data format of warehouse and line-side inventory levels

| Date  | TYP | ENGN | AXLE | PIPE | CONT | BODY | DASH | ECUT | STWL | EXST | TRIM | WIRE | SNSR | DMPR | CNGK | HYDK | SUB |
|-------|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 09/01 | CNG | 138  | 154  | 129  | 102  | 158  | 96   | 404  | 251  | 157  | 259  | 250  | 451  | 253  | 320  | -    | 306  |
| 09/01 | DM  | 63   | 66   | 67   | -    | 94   | 41   | 80   | 83   | 46   | 91   | 42   | 88   | 93   | -    | 49   | 73   |
| 09/01 | GLXi| 189  | 159  | 156  | -    | 150  | 103  | 153  | -    | 123  | 229  | 266  | 157  | 126  | -    | 100  |
| 09/01 | HYD | 46   | 60   | 72   | -    | 104  | 48   | 117  | 98   | 69   | 108  | 185  | 205  | 205  | 170  | -    | 139  |
| 09/01 | NEW | 27   | 112  | 122  | -    | 180  | 43   | 174  | -    | 119  | 164  | 185  | 285  | 219  | -    | -    | 142  |
| 09/01 | X132| 127  | 310  | 128  | -    | 137  | 246  | 256  | 220  | 181  | 250  | 255  | 263  | 268  | -    | 267  | 302  |
| 09/02 | CNG | 106  | 153  | 129  | 95   | 152  | 96   | 403  | 250  | 154  | 259  | 259  | 457  | 255  | 300  | -    | 314  |
| 09/02 | CNG-H| 65   | 62   | 65   | 57   | 75   | 50   | 124  | 126  | 156  | 81   | 167  | 182  | 130  | 236  | 219  | 110  |
| 09/02 | DM  | 63   | 64   | 64   | -    | 139  | 39   | 73   | 81   | 52   | 176  | 49   | 87   | 85   | -    | 91   | 77   |
| 09/02 | GLXi| 171  | 151  | 159  | -    | 155  | 102  | 158  | -    | 129  | 266  | 253  | 155  | 126  | -    | -    | 104  |
| 09/02 | NEW | 23   | 117  | 124  | -    | 184  | 53   | 168  | -    | 124  | 163  | 182  | 285  | 222  | -    | -    | 142  |
| 09/02 | X132| 120  | 243  | 121  | -    | 139  | 246  | 255  | 223  | 181  | 258  | 259  | 266  | 268  | -    | 271  | 300  |

ENGN engine; AXLE, rear axle; PIPE, fuel pipe set; CONT, compressed net gas container; BODY, special parts of body; DASH, dashboard or instrument panel; ECUT, electronic central unit; STWL, steering wheel; EXST, exhaust; TRIM, trim parts; WIRE, wiring set; SNSR, sensors set; DMPR, dampers set; CNGK, compressed net gas kit; HYDK, hydraulic kit; SUB, sub assembly required parts.
the probable response variable (DPC) can be calculated by the coefficients derived from logistic regression analysis.

\[
\text{logit} = 0.93153 - 0.39432 \times \text{TYP(1)} - 0.56284 \times \text{TYP(2)} - 0.7653 \times \text{TYP(3)} - 0.56226 \times \text{TYP(4)} - 2.28933 \times \text{TYP(5)} - 1.43129 \times \text{TYP(6)} - 0.00052 \times \text{INV1} - 0.00157 \times \text{INV2} + 0.21325 \times \text{CLR(1)} + 0.29959 \times \text{CLR(2)} - 1.13775 \times \text{CLR(3)} - 0.44392 \times \text{CLR(4)} + 0.01683 \times \text{DPP} - 0.00162 \times \text{BS} + 0.2752 \times \text{SEQ(1)} + 0.23753 \times \text{SEQ(2)} + 0.20142 \times \text{SEQ(3)} - 0.00016 \times \text{SRT} + 0.00042 \times \text{ACP} - 0.00047 \times \text{PBS} + 0.2752 \times \text{SEQ(1)} + 0.23753 \times \text{SEQ(2)} - 0.02589 \times \text{SEQ(3)}
\]

\[p = \frac{\exp(\text{logit})}{1 + \exp(\text{logit})}. \tag{5}\]

Table 5 summarizes the test results of null hypothesis in which all the coefficients associated with predictors equal 0. The test statistic \(G = 230.037\) with a \(p\)-value of 0.000 implies that there is at least one estimated coefficient that is different from 0. The results of Pearson, deviance, and Hosmer-Lemeshow goodness-of-fit tests have been also summarized in Table 5.

### Table 5 Goodness-of-fit tests

| Test                  | df | \(P\) value |
|-----------------------|----|-------------|
| All slopes are zero (G)| 23 | 0.000       |
| Pearson (\(\chi^2\))  | 584| 0.982       |
| Deviance (\(\chi^2\)) | 584| 0.947       |
| Hosmer-Lemeshow (\(\chi^2\)) | 8  | 0.901       |

In this study, there is insufficient evidence to claim that the LR model does not fit the data adequately because the \(P\)-values for all tests are larger than the significance level of 0.05. Therefore, the LR model shown in Equation 5 is appropriate in explaining the DPC prediction.

The association between the response variable and predicted probabilities has been evaluated by some measures such as Somers’ \(D\), Goodman-Kruskal’s gamma, and Kendall’s tau-a in our case; the summary of results is listed in Table 6. The measures indicate that there is a close correspondence between DPC and its predicted probabilities.

### Table 6 Measures of association

| Pairs      | Number | Percent | Summary measures     | \(P\) value |
|------------|--------|---------|----------------------|-------------|
| Concordant | 67,781 | 84.6    | Somers’ D            | 0.69        |
| Discordant ties | 12,163 | 15.2   | Goodman-Kruskal’s gamma | 0.70   |
|            | 151    | 0.2     | Kendall’s tau-a      | 0.30        |
| Total      | 80,095 | 100     |                       |             |
instances which have been selected from the worst to the best states.

From the perspective of daily planning accuracy, the logistic regression model correctly identified 109 of 124 observations (refer to instance 1). The accuracy of each method can be simply calculated by dividing the number of observed actual productions, which are respondents of production plans (DPC = 1), to the total number of production plans. As shown in Table 7, by the proposed DSS method, more reliable detailed production plans can be submitted than by the classic method.

**Conclusions**

This study presents an application of statistical multivariate method together with the solver module in production activity control of an Iranian automotive manufacturer and introduces a revised decision support system which can provide a productive tool for knowledge workers to offer more reliable detailed production plans.

The proposed method is based on the use of principal component analysis to reduce the extensive dimension of shop floor data and logistic regression analysis to make a predictive tool and pre-check of daily production plan capability to improve the effectiveness of decision making. In this case study, it is shown that the revised DSS works more reliably and more accurately.

For future studies, either prediction accuracy or data reduction techniques may be improved by applying other specialized models of logistic regression. Manufacturers can also further adjust the proposed prediction models to accord with their production environments and data availability.

**Competing interests**

The authors declare that they have no competing interests.

**Authors’ contributions**

SM has made substantial contributions to the concept, design, and acquisition of data and analysis and interpretation of data. MB has been involved with the main idea, drafting the manuscript, revising it critically for important intellectual content, and has given final approval of the version to be published. Both authors read and approved the final manuscript.

**Authors’ information**

SM is a PhD student at Payme Noor University. MB is an associate professor at Shahed University.

**Acknowledgments**

The authors would like to thank the Logistic and Industrial Engineering department and Production Logistics department of SAIPA Company for supporting this case study in data acquisition and serious cooperation.
Author details

1Industrial Engineering Department, Payme Noor University (PNU), Tehran, Iran. 2Industrial Engineering Department, Shahed University, Tehran, Iran.

Received: 16 June 2012 Accepted: 7 May 2013
Published: 25 May 2013

References

Aguilera AM, Escabias M, Valderrama MJ (2006) Using principal components for estimating logistic regression with high-dimensional multi collinear data. Computational Statistics & Data Analysis 50:1905–1924

Aytug H, Lawley MA, McKay K, Mohan S, Uzsoy R (2005) Executing production schedules in the face of uncertainties: a review and some future directions. European Journal of Operational Research 161(1):86–110

Bezwick V, Cheek L, Ball J (2005) Statistics review 14: logistic regression. Critical Care 9(1):112–118. doi:10.1186/cc3045

Brucker P (2007) Production control: a universal conceptual framework. Production Planning and Control 18(1):1–16

Caricato P, Greco A (2009) A DSS for production planning focused on customer service and technological aspects. Robotics and Computer-Integrated Manufacturing 25:871–878

Marefalias CT, Sung C (2009) Integration of production planning and scheduling: overview challenges and opportunities. Computers Chemical Engineering 33:1919–1930

Cunha PF, Wendahl HP (2005) Knowledge acquisition from assembly operational data using principal components analysis and cluster analysis. CIRP Annals - Manufacturing Technology 54(1):27–30

Farella RR, Maners TC (2005) A relational database approach to a linear programming-based decision support system for production planning in secondary wood product manufacturing. Decision Support Systems 40:183–196

Framinan JM, Ruiz R (2010) Architecture of manufacturing scheduling systems: literature review and an integrated proposal. European Journal of Operational Research 205:237–246

Günter HO, van Beek P (eds) (2003) Advanced planning and scheduling solutions in process industry. Springer, Berlin

Hatzikonstantinou O, Athanassou E, Pandelis DG (2012) Real-time production scheduling in a multi-grade PET resin plant under demand uncertainty. Computers and Chemical Engineering 40:191–201

Hosmer DW, Lemeshow S (2000) Applied Logistic Regression, 2nd edn. John Wiley & Sons, New York

Jahangirian M, Elfaki T, Naseer A, Stergioulas LK, Young T (2010) Simulation in manufacturing and business: a review. European Journal of Operational Research 203:1–13

Jolliffe IT (2002) Principal component analysis, 2nd edn. Springer Series in Statistics, Springer-Verlag, New York

Jourdat L, Basseur M, Talbi E (2009) Hybridizing exact methods and metaheuristics: a taxonomy. European Journal of Operational Research 199(3):620–629

Ko CH, Wang SF (2010) GA-based decision support systems for precast production planning. Automation in Construction 19(7):907–916

McKay KN, Black GI (2007) The evolution of a production planning system: a 10-year case study. Computers in Industry 58:756–771

McKay KN, Wiers VCS (2003) Integrated decision support for planning, scheduling and dispatching tasks in a focused factory. Computers in Industry 50(15–14

McKay KN, Wiers VCS (2004) Practical production control: a survival guide for planners and schedulers. J Ross Publishers, Boca Raton

Mele FD, Musulin E, Puigjaner L (2005) Supply chain monitoring: a statistical approach. Computer Aided Chemical Engineering 20:1375–1380

Mokry (2009) A decision support system for the production control of a semiconductor packaging assembly line. Expert Systems with Applications 36:4423–4424

Peidina D, Mula J, Polera R, Verdegay JL (2009) Fuzzy optimization for supply chain planning under supply, demand and process uncertainties. Fuzzy Sets and Systems 160:2640–2657

Ribas I, Leisten R, Framin JM (2010) Review and classification of hybrid flowshop scheduling problems from a production system and a solutions procedure perspective. Computers & Operations Research 37:1439–1454

Rolo M, Martinez E (2012) Agent-based modeling and simulation of an autonomous manufacturing execution system. Computers in Industry 63:53–78

Ross JW, Bernardo AL (2011) Single and parallel machine capacitated lotsizing and scheduling: new iterative MILP-based neighborhood search heuristics. Computers & Operations Research 38(12):1816–1825

Sharma S (1996) Applied Multivariate Techniques. John Wiley & Sons, New York

Simchi-Levi D, Kaminsky P, Simchi-Levi E (2000) Designing and managing the supply chain: concepts, strategies and case studies, 3rd edn. McGraw-Hill/ Irwin, Boston

Sotiris G, Athanasiou S, Ilias T (2008) A decision support system for detailed production scheduling in a Greek metal forming industry. MIBES Transactions 2(1):41–59

Torab SA, Ebadian M, Tanha R (2010) Fuzzy hierarchical production planning (with a case study). Fuzzy Sets and Systems 161(11):1511–1529

Vargas V, Metters R (2011) A master production scheduling procedure for stochastic demand and rolling planning horizons. Int J Production, Economics 132:296–302

Verdeireme PM, Floudas CA (2009) Integrated operational planning and scheduling under uncertainty. Computer Aided Chemical Engineering 26:381–386

Volling T, Spengler TS (2011) Modeling and simulation of order-driven planning policies in build-to-order automobile production. International Journal of Production Economics 131(1):183–193

Vollmann ET, Berry WL, Whybark CP, Jacobs CP (2005) Manufacturing planning and control for supply chain management, 5th edn. McGraw-Hill Education, Singapore

Yao S, Jiang Z, Li N (2012) A branch and bound algorithm for minimizing total completion time on a single batch machine with incompatible job families and dynamic arrivals. Computers & Operations 39(5):939–951

Submit your manuscript to a SpringerOpen journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com