CHEMICAL ENGINEERING TRANSACTIONS
VOL. 43, 2015
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.aidic.it/cet
Chief Editors: Sauro Pierucci, Jiří J. Klemeš

ISBN 978-88-95608-34-1; **ISSN** 2283-9216
DOI: 10.3303/CET1543223

How to Account for Market Volatility in the Conceptual Design of Chemical Processes

Davide Manca*, Alberto Conte, Riccardo Barzaghi

PSE-Lab, Process Systems Engineering Laboratory, Dipartimento di Chimica, Materiali e Ingegneria Chimica "Giulio Natta", Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy
davide.manca@polimi.it

The conventional definition of economic potentials as proposed by Douglas (1988) considers the operative expenses (OPEX) of chemical plants fixed. The main problem raised by conventional economic potentials is that they are static even if they are used for conceptual design activities. The volatility of markets coupled to demand uncertainty and fluctuation of quotations calls for an innovative approach to economic assessment, capable of removing the static attribute in favor of a dynamic approach. The paper presents a systematic approach to evaluate the possible evolution of the price of commodities to assess a set of possible economic scenarios. Different autoregressive model inputs with standard and moving average time series are analyzed to forecast the distribution of commodity prices that contributes, at different levels, to the definition of dynamic economic potentials.

1. Introduction

A systematic procedure to evaluate the feasibility study of an industrial plant was proposed by Douglas (1988) in his book on conceptual design of chemical processes. That procedure is based on the evaluation of five different economic potentials. It is mandatory for an industrial process to have each economic potential positive in order to be profitable. The feasibility study proposed by Douglas calls for the economic assessment of both CAPEX (CAPital EXpenses) and OPEX (OPerative EXpenses) terms. The assessment can be carried out for either green-field or brown-field projects. CAPEX terms are usually evaluated by using the formulas proposed by Guthrie (1969) and actualized with cost indexes. The most used cost indexes in chemical plants are: Marshall & Swift (All industries, Process industry), Nelson-Farrar (Refinery construction index), and Chemical Engineering (Plant cost index). OPEX terms are usually evaluated by assuming constant the price of commodities and utilities used to produce the final products. This steady state approach to the economic assessment of process design does not take into account the intrinsic variability of prices/costs due to: market fluctuations, demand modification, financial fluctuations, offer/demand oscillation, volatility of prices/costs, climate change, shortages, overproduction, seasonal/annual periodic variations, natural disasters, and anthropic events (e.g., floods, earthquakes, wars, political and financial crises) (Manca, 2012). Obviously, that economic assessment is rather restrictive. Aim of this paper is to define a methodology capable of taking into account the dynamic features of OPEX terms. In particular, we focus on forecasting the dynamic variation of prices and costs of commodities based on mathematical models. These models can be used to forecast the price of commodities throughout the economic life cycle of the plant, which is typically about 10-20 y.
Douglas studied the hydrodealkylation of toluene to produce benzene, so we focus on the price of these commodities. The prices of benzene and toluene not only vary continuously with an oscillating behaviour, but sometimes the price of toluene is higher than the price of benzene, making the process evidently unprofitable. The exploitation of econometric models makes more realistic and detailed the feasibility study of industrial processes. In addition, the proposed models can be used also to allocate resources in plant operations like scheduling and planning, i.e. over shorter time intervals.
Before focusing on the specific econometric models, it is worth introducing the so-called dynamic economic potentials (DEP), which lay their foundations on the original Douglas' economic potentials (EP), but feature an important difference based on the quotations dependency from time. For instance, the dynamic economic potential of second level ($DEP2$) can be formulated as follows:

Please cite this article as: Manca D., Conte A., Barzaghi R., 2015, How to account for market volatility in the conceptual design of chemical processes, Chemical Engineering Transactions, 43, 1333-1338 DOI: 10.3303/CET1543223

$$DEP_{2,k}\left[\frac{\$}{year}\right] = \frac{\sum_{i=1}^{NM}(\max(0,(\sum_{p=1}^{NP} C_{p,i,k}F_P - \sum_{r=1}^{NR} C_{r,i,k}F_R -)) \, nHPM)}{\frac{NM}{12}} \quad k = 1, \dots, NS \tag{1}$$

Where:

- NP = number of products
- NR = number of reactants
- FP = mass/molar flow rate of products
- FR = mass/molar flow rate of reactants
- CP/CR = mass/molar cost of products and reactants
- k = index of simulated scenario
- NS = number of different scenarios
- NM = number of months
- $nHPM$ = number of working hours in a month

It is worth underlining the presence of k index in Eq. (1). In fact, the real price trend includes some intrinsic stochastic elements. To consider this aspect we introduce a random component in our price forecast models, whose nature will be discussed in detail in the following Sections. By doing so, different simulations produce different results. Therefore, it is possible to obtain NS different evolution scenarios of $DEP2$ based on different dynamics of benzene and toluene prices. The DEP distribution based on NS scenarios makes the economic assessment switch from a deterministic to a stochastic solution of the problem.

2. Methods

The implementation of mathematical models for forecasting purposes requires the knowledge and the critical use of specific math tools to carry out a time-series analysis. For specific information about correlation**,** correlogram, autocorrelogram see Dagum and Cholette (2002) and Manca (2013). Other tools used are Adjusted R^2 and AIC to analyze the goodness of the models, and moving average as linear filter of time series.

- **Adjusted R^2** is used to compensate for the addition of variables to the model (Theil, Henri, 1961). As more independent variables are added to the regression model, unadjusted R^2 will generally increase but there will never be a decrease. This will occur even when the additional variables do little to help explaining the dependent variable. To compensate for this, Adjusted R^2 is corrected according to the number of independent variables in the model. The result is an Adjusted R^2 that can either increase or decrease depending on whether the addition of another variable adds or does not add any further detail to the explanatory capability of the model. Adjusted R^2 is always lower than the unadjusted one.

  $$R^2_{Adjusted} = R^2 - (1 - R^2)\frac{p}{n-p-1} \tag{2}$$

- **Akaike information criterion (AIC)** is a method to evaluate and compare different models, developed by the Japanese mathematician Hirotsugu Akaike (Akaike, 1976). AIC takes into account the goodness of fitting models and their complexity. It comes from information theory and offers a relative appraisal of the information lost when a given model is used to simulate the process that generated the data. The rule usually adopted to select the best candidate model is to prefer the one with the lowest AIC value (Chatfield, 2000).

  $$AIC = \; n\, LN\left(\frac{SSE}{n}\right) + \; 2p \tag{3}$$

  Where:
  - n number of elements in time series;
  - SSE sum of squared errors of prediction;
  - p number of parameters.

- **Simple moving average**. It is used in statistical analysis to determine the mean value of a series of elements (e.g., moving average price) at a given time. It is calculated by using the arithmetic mean of a certain number of elements of the time series vector of data.

  $$MA_t(\tau) = \frac{1}{\tau}\sum_{\vartheta=t-\tau}^{t-1} D_\vartheta \tag{4}$$

- **Central moving average**. It is a variant of the simple moving average. Equally spaced data are used on both sides of each time-series point. Formally, the central moving average of semi-amplitude $\tau/2$ is:

$$CMA_t(\tau) = \frac{1}{\tau+1} \sum_{\vartheta=t-\frac{\tau}{2}}^{t+\frac{\tau}{2}} D_\vartheta \quad (5)$$

where $\tau/2$ is an integer number. The central moving average is evaluated for a total of $\tau+1$ values: $\tau/2$ values on the right, $\tau/2$ values on the left and the central value at D_ϑ. If the damping value τ is an even number, a derived expression is used (Damiano, 2008).

2.1 Correlation between crude oil and benzene

This paper sheds some more light on modeling the commodities prices for forecasting purposes. Starting from the work of Manca (2013) and for the sake of conciseness, this article focuses just on benzene and suggests some new econometric models to forecast its price over long-term horizons. As extensively discussed in Manca (2013) and Rasello and Manca (2014), it is useful to identify a functional dependency of commodity prices (e.g., benzene) respect to the quotations of a reference component that is widely available and plays a significant role on their quotations.

Crude oil is a good candidate to play the role of reference component because it is a precursor of several derived products and specifically of benzene. The price of crude oil is broadly known and periodically updated. By analyzing the crude oil and benzene trends, one can observe a certain dependency between them. By considering that the crude oil volumes exchanged in the world are greater than the benzene ones, and that benzene is a derived product of crude oil, it seems reasonable to claim that benzene prices depend on the crude oil quotations.

A second step recommended in understanding and then building a dedicated econometric model consists in observing the functional dependency of benzene from crude oil. This can be done quantitatively by diagramming the correlogram of time series of benzene prices respect to crude oil quotations. For lower time shifts, the correlation between the time series is high. Conversely, an increase of the time shift brings to a significant reduction of the correlation, showing that there is not a strict dependency of benzene from petroleum at high time delays.

2.2 Autocorrelation of benzene

A further step to understand and build an econometric model consists in observing the self-dependency of the commodity to be identified. This can be carried out by the autocorrelogram that quantifies how previous prices of benzene afflict the present price. The autocorrelogram allows also outlining possible periodic dependencies of the commodity prices.

The correlation is obviously equal to 1 when the displacement is null, as a time series is perfectly correlated to itself. The correlation values become negligible just three months after the present quotation. This evidence allows observing that a robust econometric model should not take into account any dependency of benzene prices from values older than 3-4 months.

3. Proposed model

This Section is devoted to identify the structure and features of a family of econometric models capable of describing the functional dependency of a general commodity price (specifically the benzene one) respect to the quotation of the reference component (e.g., crude oil). A good candidate for the family of econometric models is that of autoregressive models (better known as ARX from AutoRegressive model with eXogenous input, (Ljung, 1998)) since they are linear, rather simple, and flexible as far as the number of dependent and independent terms are concerned. Generally speaking, a mixed autoregressive model with p time delays for the independent variable and q time delays for the dependent variable is defined as $ADL(p, q)$ which stands for Autoregressive Distributed Lag (Stock and Watson, 2003; Manca 2012; Manca 2013). In order to underline and exploit the strong dependency of benzene prices either from crude oil quotations or from its own prices, Eq.s (6) and (7) propose two rather simple models that depend individually on those contributions:

$$P_B(t) = A_B + B_B\, P_{CO}(t-1) \quad (6)$$

$$P_B(t) = A_B + B_B\, P_B(t-1) \quad (7)$$

where P_{CO} and P_B are, respectively, the prices of crude oil and benzene.

To determine the adaptive parameters of these models (i.e. A_B and B_B of Eq.s (6) and (7)), it is necessary to minimize Eq.s (8, 9), which are the sum of squared errors between the model P_B and real prices $P_{B,real}$. The number of sampling points used to determine the coefficients A_B, B_B are 40 (from January 2007 to April 2010, on monthly basis). We used as $P_B(t-1)$ the real benzene quotation and not the predicted one. This is the so-called *one-step forward* approach, which is useful if the aim of the models is to predict the prices of the

commodity over a short-term period. The same models can be used according to a so-called *fully predictive* approach by considering as $P_B(t-1)$ not the real, but the predicted value.

$$\min_{A_B, B_B} \sum_{t=1}^{NM} \left[(A_B + B_B P_{CO}(t-1)) - P_{B,real}(t) \right]^2 \quad (8)$$

$$\min_{A_B, B_B} \sum_{t=1}^{NM} \left[(A_B + B_B P_B(t-1)) - P_{B,real}(t) \right]^2 \quad (9)$$

The quite simple models of Eq.s (6) and (7) show a fairly good capability to reproduce the trend of benzene prices over a three-year period and suggest that a model, comprising both contributions, would perform even better. In addition, it is highly recommended to consider the crude oil term in the econometric model as it allows anticipating the influence of sudden petroleum variations on the price of benzene.
Based on the results previously achieved for both the independent and dependent variables the first candidate model is:

$$\textbf{\textit{Model M1}}: \; P_B(t) = A_B + B_B P_{CO}(t) + C_B P_B(t-1) \quad (10)$$

This model introduces a time delay, roughly estimated in one month, between the real and model data. To reduce that delay, it would be attractive to shift the model curve one month back (i.e. on the left). The back shift of the model can be formalized mathematically by means of Eq. (11). Unfortunately, the formal translation of the back shift results into an identity that makes Eq. (11) pointless.

$$P_B(t-1) = A_B + B_B \, P_{CO}(t) + C_B \, P_B(t-1) \quad (11)$$

As a matter of facts, the minimization of Eq. (8) according to problem (11) allows determining the following set of parameters $A_B = B_B = 0$ and $C_B = 1$ that transform Eq. (11) into an identity.
To reduce the residual errors between predicted and real data, it would be worth, at least ideally, to formulate a model capable of zeroing the systematic time shift that characterizes the delay of benzene prices respect to petroleum prices. Mathematically this can be achieved by introducing the crude oil quotation at time $(t+1)$. However, the ideal mathematical formulation ought to have a confrontation with real matters that cannot rely on future values of quotations used as input values in the model. To overcome this critical point, it is possible to appraise the future $(t+1)$ value of crude oil price by extrapolating the linear regression (i.e. straight line) of the two previous quotations at time t and $t-1$ (which are available on the trading markets as they are not positioned in the future). One could also think of improving the forecasting capability of the regression by introducing a non-linear trend such a parabolic term. However, by doing so, the parabolic extrapolation would introduce both complexity and possible disturbances on the overall model and would require more terms in the past thus leading to a kind of sluggish response by the econometric model.
Eq. (12) describes the proposed model that implements the future input element at $(t+1)$ based on the abovementioned linear regression:

$$\textbf{\textit{Model M2}}: \; P_B(t) = A_B + B_B \, P_{CO}(t+1) + C_B P_B(t-1) \quad (12)$$

$$P_{CO}(t+1) = m_{CO} \, (t+1) + q_{CO} \quad (13)$$

where m_{CO} and q_{CO} are, respectively, the slope and the intercept of the straight line. With the same number of parameters, we observe a modest decrease of residual errors respect to model M1. In fact, the correlation increases of 2% (see also Table 1). An alternative approach to improve the identification quality of the model consists in increasing the number of elements for either the independent or dependent variables. This increase in complexity must be counterbalanced by paying attention to avoid the risk of overparameterization (Manca, 2013) and can be steered by the correlation analysis based on both correlograms and autocorrelograms. Eq. (14) shows the model structure that better mediates complexity and detail:

$$\textbf{\textit{Model M3}}: \; P_B(t) = A_B \; + \; B_B \, P_{CO}(t) + \; C_B P_{CO}(t-1) + \; D_B P_B(t-1) + \; E_B P_B(t-2) \quad (14)$$

In this case, the correlation increases from 89.7% (Model M1) to 91.7% and the Adjusted R^2 value increases from 78.7% (Model M1) to 81.5%. The proposed models can be analyzed by using the estimators previously discussed, i.e. Adjusted R^2 and AIC. The best model is M2 for both Adjusted R^2 and AIC indicators. This means that models with a rather high number of parameters (e.g., model M3) are not justified.

3.1 Moving average models

The previous models are based on the original time series of benzene and crude oil. The errors between previous models and real prices are partially due to the impossibility to forecast the stochastic contributions that are one of the features of real quotations. These stochastic terms are characterized by rather

high-frequency fluctuations. These considerations took us to the decision of considering just a reduced number of terms to evaluate the moving average. As the proposed models are tailored to feasibility studies of industrial plants/processes, the stochastic component on the short run plays a minor role and has a reduced influence. Whoever uses these models is mainly interested in assessing the effect played by the stochastic terms throughout the so-called long-term horizon, which is primarily affected by low frequency fluctuations.
When it comes to the noise reduction of real quotations by means of filtering the high-frequency fluctuations, it is crucial to make a distinction between centered (i.e. symmetric) and asymmetric formulas for the evaluation of the moving average smoothed trend of prices/costs. The central moving average does not increase the lag between the predicted values and the original ones. On the contrary, the asymmetric moving average increases the time delay proportionally to the number of involved terms. To reduce undesired disturbances on the real trend of quotations that may be introduced by the moving average and focus on filtering just the high frequency contributions to quotations, it is advisable to use the minimum number of elements of the central formulation, i.e. three. If one used a higher number of terms in the central moving average formula (e.g., five, seven, …) then there would be the risk of running into the cancellation of major oscillations like the ones that occurred between July 2009 and January 2010, which had nothing to do with stochastic components but were representative of tangible bullish and bearish short-, medium-term intervals.
This Section focuses on the same models considered above but the adopted input data are based on suitable historical series averaged with three terms. The models identified with central-moving-average input data (i.e. mathematically filtered values) show smoother trends respect to the models identified with original unfiltered price values. It is worth observing that the models identified with filtered input data are not suitable for financial applications, which require prompt responses also based on the nervous and noisy shocks featuring high frequency stochastic contributions. On the contrary, filtered models are reliable for PSE applications that run on long-term horizons and are meant to grab the whole panorama of economic scenarios based on probabilistic distributions of feasibility studies for both conceptual design and supply chain applications (see Manca 2013; Mazzetto *et al.*, 2013; Rasello and Manca, 2014 for further details).

Table 1: Results of the analysis with AIC and Adjusted R^2 criteria (M.A. = moving average).

Model	M.A.	Correlation	AIC	Adjusted R^2
M1	No	0.8977	131	0.7876
M2	No	0.9119	126	0.8159
M3	No	0.9175	128	0.8156
M1	Yes	0.9418	117	0.8724
M2	Yes	0.9584	106	0.9081
M3	Yes	0.9847	83	0.9624

The models identified respect to real data filtered with central moving average terms show a correlation between original and predicted time series that is higher than the models identified with original (i.e. unfiltered) prices. The cancellation of high frequency stochastic fluctuations is responsible for this improvement.
As far as the central moving average models are concerned (i.e. filtered models), the best model is M3. In this case, a larger number of parameters, and so the general complexity of the model (respect to Model M2), is justified according to both AIC and Adjusted R^2 ranking criteria.
It is also worth analyzing the maximum and mean errors produced by the models (as shown in Table 2).

Table 2: Analysis of residual errors in filtered models.

Model	M.A.	Mean absolute relative error	Maximum absolute relative error	Minimum absolute relative error
M1	Yes	0.1378	1.2194	0.0010
M2	Yes	0.1216	1.0050	0.0018
M3	Yes	0.0833	0.5013	0.0001

There is quite a significant relative maximum error in every model. If data are analyzed, the higher error refers to the big economic/financial crisis of the 2008 summer-fall. This error is due to the strong fall of prices with a systematic shift of models respect to real quotations as discussed in previous Sections. In that period the average relative error is in the order of about 10%. Figure 1 shows the values obtained with Model M3 using, or not using, moving average data.

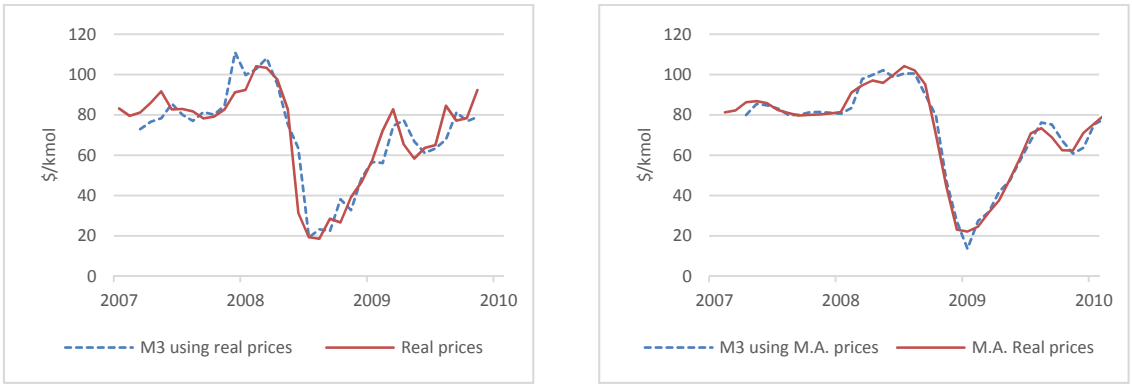

*Figure 1: Model M3 with original unfiltered input data (left) and with filtered input data (*i.e. *moving average prices based on three-month centred terms) (right) for a three-year period.*

4. Conclusions

The econometric models proposed in this paper use as reference component the crude oil. To improve the forecasting precision of such models it is therefore important to deploy models capable of predicting possible scenarios of crude oil quotations. The paper discussed and showed how, by using different tools for the time series analysis, it is possible to create econometric models to forecast the price/cost benzene which is just paradigm of the larger set of commodities. If the aim of the analysis concerns the financial field and then, the target is to forecast the so-called quotation shocks, then the models proposed in this manuscript are not appropriate. Conversely, if the interest is focused on PSE applications that cover longer time horizons, then the proposed models become valid and rather promising. For standard models, i.e. those based on original market prices/quotations, we showed that a model with three parameters (i.e. M2) and the dependency of crude oil extrapolated linearly at time $(t+1)$ is better than a five-parameter model (i.e. M3), according to both Adjusted R^2 and AIC ranking criteria. On the contrary, for central moving average models, the increment of parameters number to five is justified (i.e. model M3). As the commodities are generally subject to a considerable volatility, it is advisable to develop a model that, respecting a certain degree of simplicity, can forecast these fluctuations. On the base of the obtained data, we recommend to center average the input data (with a reduced number of elements contributing to the moving operator) so to remove the high-frequency stochastic features of real prices. These models allow switching from a deterministic to a stochastic approach for economic assessment of industrial processes, and improving the reliability of feasibility studies.

References

Akaike H., 1976, An information criterion (AIC). Math. Sci., 14, 153, 5-9.
Chatfield C., 2000, Time-series forecasting. Chapman & HALL/CRC.
Dagum E.B., Cholette P., 2006, Benchmarking, Temporal Distribution, and Reconciliation Methods for Time Series. New York, USA: Springer.
Damiano M., 2008, Demand planning. Milan, Italy, Springer Verlag.
Douglas J.M., 1988, Conceptual Design of Chemical Processes. New York: McGraw-Hill.
Guthrie K.M., 1969, Capital cost estimating. Chemical Engineering, 76, 6, 114.
Ljung L., 1998, System identification: Theory for the user. New York: Prentice Hall.
Manca D., 2012, A methodology to forecast the price of commodities. Comp. Aid. Chem. Eng., 31, 1306-1310.
Manca D., 2013. Modelling the commodity fluctuations of OPEX terms. Comp. & Chem. Eng. 57, 3-9.
Mazzetto F., Ortiz-Gutiérrez R.A., Manca D., Bezzo F., 2013, Strategic design of bioethanol supply chains including commodity market dynamics. I&ECR, 52, 10305-10316.
Rasello R., Manca D., 2014, Stochastic price/cost models for supply chain management of refineries. Computer aided Chemical Engineering, 33, 433-438.
Stock J.H., Watson M.W., 2003, Introduction to Econometrics. London: Pearson Education.
Theil H., 1961, Economic Forecasts and Policy. Holland, Amsterdam: North Holland Publishing Co.