Control-Relevant Estimation of Demand Models for Closed-Loop Control of a Production-Inventory System

Jay D. Schwartz and Daniel E. Rivera

Abstract—The development of control-oriented decision policies for inventory management in supply chains has received considerable interest in recent years, and demand modeling to supply forecasts for these policies is an important component of an effective solution to this problem. Drawing from the problem of control-relevant parameter estimation, this paper presents an approach for demand modeling in a production-inventory system that relies on a control-relevant weight to tailor the emphasis of the fit to the intended purpose of the model, which is to provide forecast signals tactical inventory management policies based on Internal Model Control. The formulation is multi-objective in nature, allowing the user to emphasize inventory variation, starts change variation, or a weighted combination. By integrating the demand modeling and inventory control problems, it is possible to obtain reduced-order demand models that exhibit superior performance. A systematic approach for generating these weights is presented and the benefits resulting from their use demonstrated on a representative production-inventory system case study.

Index Terms—demand modeling, control-relevance, Internal Model Control, production control, supply chain management

I. INTRODUCTION

A number of control-oriented approaches have recently been proposed to deal with the inventory management problems inherent in supply chains, based on Internal Model Control (IMC) and Model Predictive Control ([1], [2], [3], [4]). In these approaches, the goal is to maintain a desired net stock inventory setpoint by adjusting starts in the factory. Demand is treated as an exogenous disturbance signal that must be properly "rejected" by a sensibly-designed control system; a demand model provides disturbance forecasts which are then utilized by the control algorithms as anticipatory feedforward signals. The result of applying feedforward action based on demand forecasts is critical to the superior performance of these policies in comparison to feedback-only approaches. However, understanding how demand models should be developed for the sake of this class of supply chain management policies has not been fully examined. This paper attempts to develop such an understanding of disturbance/demand modeling and the effects of modeling error on control-oriented tactical decision policies for a production-inventory system, the basic unit of a supply chain. The relationship between demand forecast error and changes in inventory and starts are examined for a multiple degree-of-freedom combined feedback-feedforward policy based on Internal Model Control. To accomplish this goal, we draw from ideas in control-relevant parameter estimation that have been applied to problems in model reduction and system identification ([5], [6]). Results from case studies show that using this framework not only improves the performance of the closed-loop control system, but also enables planners to obtain simpler yet highly effective demand models.

Traditional approaches to modeling seek to minimize the error between prediction and actual demand in a least-squares sense, without taking into account the requirements of the intended application. In this work the performance requirements for the end-use inventory control problem are used to determine the frequency regimes over which improving the goodness-of-fit in the demand model has the most impact. The result of this analysis is a closed-form expression for a weighting function capturing essential aspects of the production-inventory problem which can then be used to define important design variables in the demand modeling problem in a control-relevant manner. For example, this weight can be used to define data prefilters [7] or, as examined in this paper, the weight function in a control-relevant frequency response curvefitting problem that fits an empirical transfer function estimate to a parametric model, or reduces a high-order demand model to a simpler, restricted complexity form.

The paper is organized as follows. A fluid analogy model for the production-inventory system and the corresponding IMC-based decision policy are developed in Section II. In Section III we present a control-relevant modeling approach for an Internal Model Control-based decision policy. The formulation is multi-objective in nature, allowing users to emphasize variations in inventory, starts changes, or their weighted combination; furthermore, the closed-form nature of the IMC policy allows the control-relevant weighting function to be determined explicitly. A case study is presented in Section IV which consists of control-relevant model reduction of a high-order demand model. Summary and conclusions are included in Section V.

II. BACKGROUND AND PROBLEM DEFINITION

A. Fluid analogy for production/inventory systems

A fluid analogy for a standard single-product production-inventory system, the simplest unit in a supply chain, is shown in Figure 1. Fluid analogies represent meaningful descriptions of supply chains associated with high volume manufacturing problems at sufficiently long time scales (for
instance, in daily or weekly decision-making). This applies to discrete-parts manufacturing problems such as semiconductor manufacturing ([1], [2], [3]). The output of a factory is stored in a warehouse where it awaits shipment to customers (retailers or distributors). The warehouse serves as a buffer in the presence of stochastic, uncertain customer demand and factory output.

The factory is modeled as a pipe with a specific throughput time \( \theta \) and yield \( K \). Inventory is modeled as material (fluid) in a tank. Delivery from the warehouse is modeled as a pipe with a transportation time \( \theta_d \). Applying the principle of conservation of mass to this system leads to a differential equation relating net stock (material inventory, \( y(t) \)) to factory starts (input pipe flow, \( u(t) \)) and customer demand (output tank flow, \( d(t) \)); this is represented by the equation

\[
\frac{dy}{dt} = K u(t - \theta) - d(t)
\]

with the corresponding transfer functions

\[
y(s) = \frac{K e^{-\theta s}}{s} u(s) - \frac{1}{s} d(s)
\]

(2)

Customer demand \( d(t) \) is treated as a disturbance signal at the tank outlet, and consists of the sum of forecasted demand \( d_F(t) \), known \( \theta_F \) days ahead of time, forecast error \( d^e_F(t) \), and unforecasted demand \( d_U(t) \), as shown below

\[
d(t) = d_F(t - \theta_F) + d^e_F(t - \theta_F) + d_U(t)
\]

(3)

\( d^e_F \) represents an ideal forecast. The demand forecast \( d_F \) is generated from a demand model \( \hat{P}_f \)

\[
d_F(s) = \hat{P}_f(s) f_i(s)
\]

which is a function of a disturbance input \( f_i \). \( f_i \) represents any measurement signal that can be used as an input to predict future demand, for example, an economic indicator. The demand forecast error \( d^e_F \) represents the signal that we wish to keep at a minimum,

\[
d^e_F(s) = d^{ideal}_F(s) - d_F(s) = E_f(s) f_i(s)
\]

(5)

with \( E_f = P_f - \hat{P}_f \) representing the demand modeling error that should be reduced in a control-relevant manner. The dynamical system model capturing the production-inventory system is defined by the equations:

\[
y(s) = \frac{P(s)}{s} u(s) - \frac{P_{d1}(s)}{s} \left( \hat{P}_f(s) f_i(s) + d^e_F(s) \right) - \frac{P_{d2}(s)}{s} d_U(s)
\]

\[
y(s) = \frac{K e^{-\theta_F s}}{s} u(s) - \frac{e^{-\theta_F s}}{s} \hat{P}_f(s) f_i(s) - \frac{e^{-\theta_F s}}{s} d^e_F(s)
\]

\[
y(s) = \frac{1}{s} d_U(s)
\]

(6)  

(7)

The model according to Equation 7 is the nominal plant model for the Internal Model Control-based tactical decision policy described in Section II.B.

B. Multi-Degree-of-Freedom Combined Feedback-Feedforward Internal Model Control

We consider a multi-degree-of-freedom combined feedback-feedforward Internal Model Control (IMC) structure [8] as a decision policy. The treatment follows as shown in [2] and [9]. With this structure independent controllers can be utilized for setpoint tracking (i.e., meeting an inventory target), measured disturbance rejection (i.e., meeting forecasted demand), and unmeasured disturbance...
rejection (i.e., satisfying unforecasted demand). Figure 2 shows the structure schematically. The controllers correspond to $Q_c$ for setpoint tracking, $Q_f$ for measured disturbance rejection, and $Q_d$ for unmeasured disturbance rejection. The controllers are designed for nominal optimal performance, then augmented with filters to reduce aggressive manipulated variable action associated with the optimal controller or to satisfy robust performance. The controllers obtained from applying this procedure to the models according to (2) and (3) are shown as follows:

**Setpoint Tracking.** The setpoint tracking mode of this control system is designed using an $H_2$-optimal controller for a step change, augmented with a first-order filter. The controller guarantees no offset for Type-1 (step) setpoint changes in the control system. The mode allows the controller to adjust safety stock inventory targets to any user-desired level.

$$Q_r(s) = \frac{s}{K(\lambda_r s + 1)^{n_r}}. \tag{8}$$

**Unmeasured Disturbance Rejection.** This mode of the control system allows the user to specify the system response to unforecasted demand changes. The design procedure relies on an $H_2$-optimal controller for ramp disturbance changes, with a generalized Type-2 filter guaranteeing no offset for both asymptotically step and ramp disturbances,

$$Q_d(s) = \frac{s(\theta s + 1)}{K(\lambda_d s + 1)^{n_d}}. \tag{9}$$

**Measured Disturbance Rejection.** The measured disturbance rejection mode of the control system performs feedforward control action by relying on a $\theta_F$-day ahead forecast signal to manipulate factory starts. The IMC controller form is defined as follows [10]

$$Q_f(s) = Q'_f(s)f_F(s) \tag{10}$$

where $Q'_f(s)$ consists of

$$Q'_f(s) = \frac{e^{-(\theta_F - \theta_d)\theta}}{K}. \tag{11}$$

if the forecast horizon is longer than the sum of the factory throughput time and delivery time ($\theta_F \geq (\theta + \theta_d)$). If the forecast horizon is shorter ($\theta_F \leq (\theta + \theta_d)$) then $Q'_f(s)$ consists of

$$Q'_f(s) = \frac{((\theta + \theta_d - \theta_F)s + 1)}{K}. \tag{12}$$

The generalized Type-2 filter $f_F(s)$ is defined as

$$f_F(s) = \frac{(n_F \lambda_F s + 1)}{(\lambda_F s + 1)^{n_F}}. \tag{13}$$

Each controller is required to be stable and proper, thus imposing the restriction that all values of the user adjustable parameter $\lambda$ be positive ($\lambda_i > 0$) and that the filter order is chosen to ensure transfer function properness ($n_r \geq 1$, $n_d \geq 3$, $n_F \geq 2$). It is appropriate to restrict design to use semiproper forms for $Q_r$ and $Q_d$ and the strictly proper case for $Q_f$. A set of representative responses is shown in [9].

It is important to understand how the IMC decision policy responds to the presence of forecasted demand error signals $d_F$. One means is to generate the frequency responses relating $d_F$ to changes in inventory $y$ and factory starts changes $\Delta u$; amplitude ratios for these frequency responses are depicted in Figure 3 for a representative set of model and controller tuning parameters. The result is a “notch” filter where forecasted demand error is attenuated at the low and high frequency regimes, but amplified in an intermediate bandwidth. The width of this bandwidth is determined by plant characteristics and controller tuning parameters. The result is a “notch” filter where forecasted demand error is attenuated at the low and high frequency regimes, but amplified in an intermediate bandwidth. The width of this bandwidth is determined by plant characteristics and controller tuning parameters. The result is a “notch” filter where forecasted demand error is attenuated at the low and high frequency regimes, but amplified in an intermediate bandwidth. The width of this bandwidth is determined by plant characteristics and controller tuning parameters.

![Figure 3: Amplitude ratios of the transfer functions relating demand modeling error $E_f$ to changes in inventory $y$ (solid blue line) and factory starts changes $\Delta u$ (dashed green line) for an IMC-based tactical decision policy. ($K = 1$, $\theta = 5$, $\theta_F = 10$, $\lambda_d = 2$, $n_d = 4$, $\lambda_F = 3$, $n_F = 2$)](image)

**III. Control-Relevant Demand Modeling Using Frequency-Weighted Curvefitting**

Formally defining a control-relevant parameter estimation problem for demand modeling requires understanding how demand modeling error $E_f$ affects the closed-loop control error $e_c = y - r$ and factory starts changes $\Delta u$. Applying the generalized plant framework as shown in Figure 4 leads to a Linear Fractional Transformation (LFT) that characterizes
this input-output relationship.
\[
\begin{bmatrix}
    e_c \\
    \Delta u
\end{bmatrix} = \left( G_{22} + G_{21} (I - G_{11} E_F)^{-1} E_F G_{12} \right) f_i
\] (14)

Equation 14 represents an explicit relationship relating the forecast model inputs \( f_i \) to the control error \( e_c \), and the factory starts change \( \Delta u \). The closed-form expression quantifies the effect of forecast error on inventory deviation and factory thrash. When applied to the IMC feedback-feedforward system shown in Figure 2 the resulting expressions are
\[
G_{11}(s) = 0
\] (15)
\[
G_{22}(s) = \begin{bmatrix} (PQ_f - Q_d) \hat{P}_f & \Psi Q_f \hat{P}_f \\ \Psi Q_d \hat{P}_f & \Psi Q_d P_d \end{bmatrix}
\] (16)
\[
G_{21}(s) = \begin{bmatrix} (PQ_d - 1) P_d \\ \Psi Q_d \Psi Q_d P_d \end{bmatrix}
\] (17)
\[
G_{12}(s) = 1
\] (18)
yielding the following LFT
\[
\begin{bmatrix}
    e_c \\
    \Delta u
\end{bmatrix} = \left( \begin{bmatrix} (PQ_f - Q_d) \hat{P}_f & \Psi Q_f \hat{P}_f \\ \Psi Q_d \hat{P}_f & \Psi Q_d P_d \end{bmatrix} \right) f_i
\] (19)
As a result of \( G_{11} = 0 \), Equation 19 is affine in \( E_f \).

In practice it is desirable to minimize the effects of forecast error on a weighted combination of the control error and starts change signals. The first step is to define a weighting function \( W^* \)
\[
W^* = \begin{bmatrix} \sqrt{1 - \gamma} & 0 \\ 0 & \sqrt{\lambda} \end{bmatrix}
\] (20)
where the parameter \( \lambda \) is used to normalize the variance of the inventory deviation and factory starts change signals, while the parameter \( \gamma \) is a user-adjustable parameter that allows a supply chain planner to estimate models to reduce inventory deviation from setpoint (\( \gamma = 0 \)), factory starts change variance (\( \gamma = 1 \)), or some combination achieved via intermediate values of \( \gamma \). We then seek to minimize the 2-norm of the weighted control error and starts change signals.

Applying the Triangle Inequality
\[
\left\| \begin{bmatrix}
    e_c \\
    \Delta u
\end{bmatrix} \right\|_2 \leq \left\| \begin{bmatrix} (PQ_f - Q_d) \hat{P}_f & \Psi Q_f \hat{P}_f \\ \Psi Q_d \hat{P}_f & \Psi Q_d P_d \end{bmatrix} \right\| \left\| f_i \right\|_2
\] (21)
allows the contribution of the demand modeling error \( E_f \) to be isolated into its own term. The control-relevant parameter estimation problem consists then of minimizing
\[
\min_{\hat{P}_f} \left\| W^* \begin{bmatrix} (PQ_d - 1) P_d \\ \Psi Q_d \Psi Q_d P_d \end{bmatrix} E_f f_i \right\|_2
\] (22)
For the scalar case the 2-norm of the combined weight leads to the weighted parameter estimation problem
\[
\min_{\hat{P}_f} \| W(\omega) E_f \|_2
\] (23)
from which the control-relevant weight \( W(\omega) \) is obtained:
\[
W(\omega) = (1 - \gamma) \left| (PQ_d - 1) P_d \right|^2 |f_i|^2 + \lambda \gamma \left| \Psi Q_d P_d \right|^2 |f_i|^2
\] (24)
From (24) it is possible to explicitly identify the factors that determine how the closed-loop system’s response impact demand modeling requirements, and consequently forecast error. These factors are the weighting parameter \( \gamma \), the normalizing parameter \( \lambda \), the dynamics of the production/inventory system \( P \), the dynamics of the integrating system \( P_d \), the tuning and structure of the feedback controller \( Q_d \), and the input to the demand model \( f_i \). It is important to note that as a consequence of the multi-degree-of-freedom formulation, only \( Q_d \) affects the closed-loop system’s response to forecast error; neither the tuning of the setpoint controller \( Q_s \) or the feedforward controller \( Q_f \) play a role.

Numerically (23) can be solved via frequency-weighted curvefitting techniques, relying on the availability of a frequency response for the true demand model. In a system identification setting, the true frequency response can arise from an empirical transfer function estimate (ETF) generated from data. The control-relevant weight (24) is applied during the model reduction process to enable a low-order demand model \( \tilde{P}_f \) that is capable of meeting two supply chain objectives: minimizing inventory deviations from setpoint, factory starts change variance, or a weighted combination. It is not necessary to obtain a parametric form for \( W(\omega) \); one can readily apply curvefitting routines such as \texttt{invfreqs} in MATLAB® to implement this estimation procedure.

IV. Case Study

In this section we show the results of applying the control-relevant demand model reduction procedure to the production/inventory system described in Section 2. For the problem considered here there is no yield loss (\( K = 1 \)), the throughput time \( \theta \) of the production node is 5 days, the forecast horizon \( \theta_f \) is 10 days. The feedback-feedforward IMC controller presented in Section 2 acts as a tactical decision policy where the controller parameters are \( \lambda_d = 2 \), \( \lambda_F = 3 \), \( n_d = 4 \), and \( n_F = 2 \). The true demand model is the high-order system defined in Equation 25.
\[
d_{dideal}(s) = \frac{0.41 s^3 + 7.54 s^8 + 39.93 s^7 + 140.2 s^6 + 459.8 s^5}{s^3 + 20 s^2 + 92 s + 894.4 s^4 + 1488 s^5} \cdot 645.8 s^4 + 855 s^3 + 839.7 s^2 + 381.4 s + 312.3 \cdot 5318 s^4 + 4518 s^3 + 4019 s^2 + 1126 s + 307.2
\] (25)
The amplitude ratio for this model is shown in Figure 5. The signal \( f_i(s) \) used for the simulations is a pulse according to:
\[
f_i(s) = 10 \left( \frac{1 - e^{-0.1 s}}{s} \right)
\] (26)
The model reduction process involves a frequency-domain curve-fit to a continuous transfer function \( \tilde{P}_f(s) \) that serves as a control-relevant demand model
\[
\tilde{P}_f(s) = K_{fit} \left( \frac{s + 0.05 f_i}{s - 0.05 f_i} \right)
\] (27)
where $K_{fit}$ is the gain, $z_{fit}$ is the zero, and $p_{fit}$ is the pole of the estimated demand model. The control-relevant weighting functions were obtained using the analysis in Section III and are shown in Figure 6. The control-relevant weights shown in Figure 6(a) correspond to values of $\gamma$ equal to 0, 0.5, and 1. All three weights act to minimize error within an intermediate bandwidth. The $\gamma = 0$ weight places greater emphasis on the lower frequencies, the $\gamma = 1$ weight gives greater emphasis to the higher frequencies, while the $\gamma = 0.5$ weight acts as a combination of the other two weights.

The results from applying the procedure are captured in Figures 6 and 7 and Table I. The relative emphasis of the control-relevant weight $W(\omega)$ (as shown in Figure 6(a)) is reflected in Figure 6(b) in terms of the frequency regions where $E_f$ is minimized. Figure 6(b) shows the demand modeling error amplitude when no control-relevant weighting is performed and when the various control-relevant weights are applied. The unweighted case displays the lowest overall error at low frequencies ($\omega \leq 0.15$ rad/day). Control-relevant weighting, however, lowers the error over a bandwidth denoted by $W(\omega)$. The $\gamma = 0$ weight yields the lowest error in an intermediate bandwidth between 0.15 and 0.4 radians/day. The $\gamma = 1$ weight yields the lowest error in an intermediate bandwidth between 0.4 and 2 radians/day. The error for the $\gamma = 0.5$ case is in between, consistent with its mixed objective.

Figure 7 shows closed-loop responses resulting from the application of the nominal approach (no weighting) and the control-relevant formulation. Table I summarizes the open- and closed-loop metrics shown in Figure 7. No control-relevant weighting leads to the best “open-loop” forecast in terms of the metric $\int (\hat{d}_p^2 - d_F^2) \, dt = \int (\hat{d}_p^2) \, dt$. However, it displays the worst result with respect to the closed-loop metric for inventory deviation, $\int (y - r)^2 \, dt = \int e_r^2 \, dt$. The closed-loop metrics are improved by the use of the control-relevant formulation (where $\gamma = 0, 1$, or 0.5). Treating the unweighted model reduction as a baseline, the control-relevant modeling procedure yields a 39.2% reduction in the inventory deviation metric when the user-adjustable parameter $\gamma$ is set to 0. The procedure yields an 8.0% reduction in factory starts changes by setting $\gamma$ equal to 1. The weighted combination ($\gamma = 0.5$) leads to a substantial improvement in $\int e_r^2 \, dt$ (27.3% decrease) with a modest gain in factory starts changes (18.8% increase). These results show that the methodology is flexible enough to accommodate user preferences, and they confirm the importance of control-relevance.

V. Conclusions

Analysis of control-oriented decision policies for production-inventory control shows that these systems are most responsive to demand modeling error in an intermediate frequency bandwidth. We have derived in this paper a closed-form expression for the impact of controller performance on modeling error for the case of...
TABLE I: Closed-loop vs. open-loop metrics summary for responses shown in Figure 7.

<table>
<thead>
<tr>
<th>Weight Type</th>
<th>$\int e^2 dt$</th>
<th>$\int (\frac{de}{dt})^2 dt$</th>
<th>$\int (\frac{d^2 e}{dt^2})^2 dt$</th>
<th>Figure</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>3.348</td>
<td>0.414</td>
<td>2.375</td>
<td>7(a)</td>
</tr>
<tr>
<td>$\gamma = 0.0$</td>
<td>2.034</td>
<td>1.132</td>
<td>6.093</td>
<td>7(b)</td>
</tr>
<tr>
<td>$\gamma = 1.0$</td>
<td>2.909</td>
<td>0.381</td>
<td>2.952</td>
<td>7(c)</td>
</tr>
<tr>
<td>$\gamma = 0.5$</td>
<td>2.435</td>
<td>0.492</td>
<td>3.607</td>
<td>7(d)</td>
</tr>
</tbody>
</table>

VI. Acknowledgments

The authors would like to acknowledge support from the National Science Foundation (CMMI-0432439) and the Intel Research Council.

References


![Fig. 7: Case Study. Closed-loop results for the unweighted (a) and weighted (b: $\gamma = 0.0$, c: $\gamma = 1.0$, d: $\gamma = 0.5$) control-relevant model reduction procedure. ($K = 1$, $\theta = 5$, $\theta_F = 10$, $\lambda_d = 2$, $n_d = 4$, $\lambda_F = 3$, $n_F = 2$)](image-url)