Disentangling Production Smoothing from the Bullwhip Effect

Robert L. Bray and Haim Mendelson

Kellogg School of Management

Stanford Graduate School of Business

November 15, 2013

Abstract

Until now, production smoothing and the bullwhip effect have shared a common measure: the difference between production variability and demand variability. This metric confounds the two effects, however, suggesting that firms that exhibit the bullwhip effect cannot smooth production. We develop new production smoothing measures that are robust to the bullwhip effect. We derive these measures from a structural econometric production scheduling model, based on demand signal processing. Applying these measures to monthly auto industry data, we find that auto manufacturers significantly smooth production, despite exhibiting a strong bullwhip effect. Further, we find that auto manufacturers actively smooth both production variability, which reflects all production fluctuations, and production uncertainty, which reflects only surprising production fluctuations.[1]

Keywords: structural estimation; production smoothing; automotive industry; demand signal processing; Generalized Order-Up-To Policy; Martingale Model of Forecast Evolution.

1 Introduction

We do three things in this article: First, we devise new production smoothing measures that are robust to the bullwhip effect. The measures suggest auto manufacturers substantially smooth production, despite production being more variable than demand. Second, we show that auto manufacturers smooth both production variability and production uncertainty. And third, we develop a new empirical approach to studying inventories, based on demand signal processing.

1.1 Production Smoothing and the Bullwhip Effect

Production smoothing is a central theme in supply chain management (Anand and Mendelson, 1997; Simchi-Levi and Zhao, 2003; Aviv, 2007; Cachon et al., 2007; Veeraraghavan and Scheller-Wolf, 2008; Boute and Van Mieghem, 2011) and production scheduling (Holt et al., 1960; Beckmann, 1961; Lippman et al., 1967;

[1]We thank Gad Allon, Achal Bassamboo, Gerard Cachon, Alex Frankel, Karan Girotra, Varun Gupta, Itai Gurvich, Nitish Jain, Jan Van Mieghem, and Alejandro Molnar for their helpful suggestions.

Kleindorfer et al., 1975; Graves, 1981; Graves et al., 1998). Firms smooth production by using inventories to absorb demand shocks in order to decrease production instability costs (Blinder and Maccini, 1991, p. 78). For example, when production cost functions are convex, due to overtime fees or capacity constraints, deviations from mean production levels are costly.

Traditionally, economists have tested for production smoothing by measuring whether demand variability (DV) exceeds production variability (PV). Yet, they have largely failed to find production smoothing, since "production is more variable than sales in all major sectors and in most industries" (Blinder and Maccini, 1991, p. 80). Economists sought to reconcile this apparent anomaly, citing measurement errors (Fair, 1989), aggregation biases (Lai, 1991), cost shocks (Miron and Zeldes, 1988; Eichenbaum, 1989), demand autocorrelations (Kahn, 1987), and non-convex production cost functions (Ramey, 1991; Bresnahan and Ramey, 1994). Nevertheless, OM researchers later explained that it's reasonable for PV to exceed DV, because of the bullwhip effect (Lee et al., 1997).

The bullwhip effect states that demand variability tends to increase up supply chains. For example, beer sales generally grow more volatile from liquor stores to breweries, as transfer batch sizes grow from six-packs to truckloads. Since its inception, the bullwhip effect has shared a common measure with production smoothing: the difference between PV and DV. Yet the phenomena are distinct: The bullwhip effect states that supply chains amplify demand variations, whereas production smoothing states that firms avoid production instability costs. Since the bullwhip pertains to upstream and downstream variabilities, it's natural to measure it with PV and DV (see Chen and Lee, 2012). However, DV has little relation to production smoothing, so the bullwhip generally drowns out the phenomenon in the traditional PV-DV measure. Accordingly, we propose a more targeted production smoothing metric: Rather than benchmark PV to DV, we benchmark PV to what PV theoretically would be in the absence of production instability costs.

Until now, it's seemed a firm could either bullwhip, with PV exceeding DV, or smooth production, with DV exceeding PV, but not both. But, this is a false dichotomy: Our new smoothing measure shows auto manufacturers both bullwhip—the median PV exceeds the median DV by 43%—and smooth production—the median indifferent-to-production-stability PV exceeds the median actual PV by 67%. We can embrace both phenomena.

1.2 Variability vs. Uncertainty

Whereas all production fluctuations contribute to production variability (PV), only surprising production fluctuations contribute to production uncertainty (PU). Denoting the unpredictability of the production process, PU reflects the *information lead time* of production forecast revisions. Until now, the production smoothing literature has only considered dampening PV. We hypothesize, however, that firms also explicitly dampen PU. That is, we hypothesize that firms smooth unpredictable fluctuations more than predictable fluctuations (Graves et al., 1998; Aviv, 2007; Chen and Lee, 2009).

To define our PU measure, we couch our production scheduling model in a supply chain setting. Following Chen and Lee (2009, 2012), we then measure PU with the implied variance of supplier inventories. And we measure *PU smoothing* by comparing what PU actually is to what it theoretically would be in the absence of production instability costs. Confirming our hypothesis, we find strong PU smoothing, with the median indifferent-to-production-stability PU exceeding the median actual PU by 155%—auto manufacturers do apparently distinguish between predictable and unpredictable production fluctuations. Moreover, we find the degree of PU smoothing increased significantly from the 1990s to the 2000s—auto manufacturers have apparently gotten better at attenuating costly production fluctuations.

1.3 Harnessing Demand Signal Processing with Structural Estimation

Calculating our production smoothing measures requires anticipating how auto manufacturers would behave if they had no impetus to stabilize production—i.e., if PV and PU were costless. Structural estimation enables researchers to conduct such counterfactual analyses (Lucas, 1976; Reiss and Wolak, 2007; Keane, 2010). The approach supposes a fully specified economic model generates a given dataset; in turn, reverse-engineering the model's primitives enables empiricists to simulate how firms would behave under various counterfactual scenarios (e.g. see Cohen et al., 2003; Olivares et al., 2008; Allon et al., 2011; Aksin et al., 2012). To estimate the degree of production smoothing, we develop a structural econometric production scheduling model.

Our structural model incorporates the *Generalized Order-Up-To Policy* (GOUTP) of Graves et al. (1998) and Chen and Lee (2009). The GOUTP "enables us to sharply delineate between the effect of order variability and the effect of order uncertainty" (Chen and Lee, 2009, p. 795) via *demand signal processing* (DSP), the transformation of demand forecast revisions into production forecast revisions. From a signal processing perspective, firms have two levers to smooth production: *signal mixing* and *signal delaying*. The former pools risk across time, which lowers PV; the latter increases information lead times, which lowers PU. We use our model to transform signal mixing and delaying patterns into PV and PU aversion estimates, from which we estimate PV and PU smoothing.

The GOUTP is "an elegant model of ... production and inventory planning" (Aviv, 2007, p. 778). Capitalizing on the GOUTP's linear form, we develop a closed-form estimator of the impulse response functions that characterize this inventory policy. The GOUTP is general and parsimonious, and thus so is our empirical framework: Our model-cum-estimators constitute a basic empirical inventory framework. Moreover, this framework underpins a new empirical approach: deriving operational insights from DSP signatures.

2 Model

Demand signal processing (DSP), the translation of demand forecast revisions into production forecast revisions, drives our model dynamics (Lee et al., 1997, p. 549). We apportion flexibility to the DSP structure, because that is the conduit through which our data speak. The Martingale Model of Forecast Evolution (MMFE) underpins our model's information structure (Hausman, 1969; Heath and Jackson, 1994), and the Generalized Order-Up-To Policy (GOUTP) governs its logistics (Graves et al., 1998; Chen and Lee, 2009). We abstract away seasonality, since our estimators wash out seasonal variations.[2] Exhibit 1 provides a glossary of our model variables.

[2]It's difficult to anticipate how seasonal production patterns would change in the absence of production instability costs, so we suppose production seasonality remains fixed under the no-impetus-to-smooth counterfactual scenario. Allowing production seasonality to depend on the cost of production volatility would only increase our already-large smoothing estimates. Note, PV exceeds DV in our auto manufacturing sample, regardless of whether or not we include seasonality.

2.1 Supply Chain

A single-firm production scheduling problem underpins our model. To motivate our PU measure, however, we initially nest this single-firm problem in a dual-firm supply chain, adding a supplier. Following Chen and Lee (2009, 2012, p. 786, 778), we then measure PU with the implied variance of the supplier's inventories. Since the supplier simply provides context for our PU measure, we select a simple supplier specification. Our final empirical specification does not include the supplier.

We consider a manufacturer that produces a single output from a single input, sourced from a single supplier. Producing one unit of output requires one unit of input. The manufacturer has production lead time $\phi \geq 0$, and the supplier has production lead time $\psi > 0$. The manufacturer backlogs demand when it stocks out, and the supplier borrows from a third party when it stocks out: the supply chain is decoupled (Kahn, 1987; Gavirneni et al., 1999; Lee et al., 2000; Chen and Lee, 2009).[3] The firms fully share information, so they also share forecasts. The manufacturer places an order each period; free to return stock, it can order up to any level (Lee et al., 1997; Aviv, 2003; Chen and Lee, 2009). It's order quantity corresponds to its production quantity; the manufacturer starts production as soon as its inputs arrive. The manufacturer eventually fulfills all demand, and stores finished-goods inventories until they sell. We ignore prices because "automakers only modestly respond with changes in price when faced with a demand shock to a particular vehicle" (Copeland and Hall, 2005, p. 233).

2.2 Demand, Production, and Inventory

An MMFE governs the manufacturer's demand (see Appendix A):

$$d_t = \mu + \sum_{l=0}^{H-1} e_l' \epsilon_{t-l}, \tag{1}$$

where d_t is the period t demand, μ is the mean demand, e_l is a vector with a one in its $(l+1)^{th}$ position and zeros elsewhere, H is the length of the forecast horizon, and $\epsilon_t = E_t[d_t, \cdots, d_{t+H-1}]' - E_{t-1}[d_t, \cdots, d_{t+H-1}]'$ is an exogenous demand signal vector observed in period t. Signal vector ϵ_t has mean zero and full rank covariance matrix $\Sigma = E[\epsilon_t \epsilon_t{}']$. By definition, ϵ_t is uncorrelated with ϵ_τ for $t \neq \tau$. Since, MMFE signals correspond to forecast revisions, we use the terms "signal" and "forecast revision" interchangeably.

Next, an extension of the GOUTP governs the manufacturer's production policy. The GOUTP restricts production to follow a linear, time-invariant function of observed demand signals. Supposing the firm's production scheduling horizon is H periods long (see Graves et al., 1998, p. 37), the GOUTP takes the form: $o_t = \mu + \sum_{l=0}^{H-1} e_l' A \epsilon_{t-l}$, where o_t is the period t production quantity, and A is an $H \times H$ signal processing matrix. Matrix A characterizes the firm's inventory policy; its columns constitute the impulse response functions (IRFs) that map demand signals into production quantities, so we refer to A as the manufacturer's IRF matrix.

Under the traditional GOUTP specification, production quantities only depend on observed demand signals. But, in practice, production depends on additional factors, such as labor availability, process yields, tool downtimes, and transfer batch sizes. To account for these auxiliary production shifters, we add exogenous error term n_t to the production quantity. We suppose the firm forecasts all factors that influence its

[3] Supplier stockouts are rare in our empirical setting, as we explain in §5.1.

production, which means the n_t errors also follow an MMFE (see Appendix A): $n_t = \sum_{l=0}^{H-1} e_l' \eta_{t-l}$, where $\eta_t = E_t[n_t, \cdots, n_{t+H-1}]' - E_{t-1}[n_t, \cdots, n_{t+H-1}]'$ is an exogenous noise signal vector observed in period t, with zero mean and covariance matrix $\Lambda = E[\eta_t \eta_t']$. By definition, η_t is uncorrelated with η_τ and ϵ_τ, for $t \neq \tau$. However, contemporaneous demand and noise signals have general covariance matrix $\Gamma = E[\epsilon_t \eta_t']$ (e.g., a macroeconomic shock can simultaneously affect demand and production).

Incorporating this n_t error term into the GOUTP yields the following period t production quantity:

$$o_t = n_t + \mu + \sum_{l=0}^{H-1} e_l' A \epsilon_{t-l} = \mu + \sum_{l=0}^{H-1} e_l' \epsilon_{t-l}^o, \qquad \text{where} \qquad \epsilon_t^o = A\epsilon_t + \eta_t. \tag{2}$$

Upon this production process, we impose Graves et al. (1998, p. 37)'s market clearing constraint that cumulative production forecasts match cumulative demand forecasts: $E_t[o_t + \cdots + o_{t+H}] = E_t[d_t + \cdots + d_{t+H}]$. This regularity condition implies $\iota' A = \iota'$ and $\iota' \Lambda = 0$, where $\iota = \sum_{l=0}^{H-1} e_l$ is a vector of ones.

Since the manufacturer sets its order quantities equal to its production quantities, and shares its forecasts with the supplier, the supplier observes its demands evolving according to (2)'s MMFE, with ϵ_t^o doubling as its period t demand signal vector (see Chen and Lee, 2009). The supplier also follows a GOUTP, with IRF matrix A_s; this IRF matrix also satisfies Graves et al. (1998, p. 37)'s market clearing constraint: $\iota' A_s = \iota'$. The supplier's period t order quantity is:

$$o_t^s = \mu + \sum_{l=0}^{H-1} e_l' \epsilon_{t-l}^{o,s}, \qquad \text{where} \qquad \epsilon_t^{o,s} = A_s \epsilon_t^o. \tag{3}$$

Now we define three operators: (i) cumulative sum operator C_x is an $(H+x) \times (H+x)$ lower triangular matrix of ones, (ii) x period delay operator D_x is a matrix equal to $(I - C_x^{-1})^x$, and (iii) matrix reshape operator I_x is an $(H+x) \times H$ version of the identity matrix. With these operators, we find expressions (1) and (2) yield the following manufacturer inventory level, at the end of period t:

$$i_t = \mu_i + \sum_{l=0}^{H+\phi-1} e_l' \epsilon_{t-l}^i, \qquad \text{where} \qquad \epsilon_t^i = C_\phi (D_\phi I_\phi A - I_\phi)\epsilon_t + C_\phi D_\phi I_\phi \eta_t. \tag{4}$$

And we find expressions (2) and (3) yield the following supplier inventory level, at the end of period t:

$$i_t^s = \mu_s + \sum_{l=0}^{H+\psi-1} e_l' \epsilon_{t-l}^{i,s}, \qquad \text{where} \qquad \epsilon_t^{i,s} = C_\psi (D_\psi I_\psi A_s - I_\psi)\epsilon_t^o. \tag{5}$$

Parameters μ_i and μ_s, in (4) and (5), are the manufacturer's and supplier's mean inventory levels, respectively.

2.3 Objectives

Expressions (1)-(5) characterize demand signal processing (DSP): The manufacturer transforms signal vectors ϵ_t and η_t into signal vectors ϵ_t^o and ϵ_t^i, with decision variable A. And the supplier transforms signal vector ϵ_t^o into signal vectors $\epsilon_t^{o,s}$ and $\epsilon_t^{i,s}$, with decision variable A_s. In turn, signal vectors ϵ_t^o, ϵ_t^i, $\epsilon_t^{o,s}$, and $\epsilon_t^{i,s}$ determine the evolution of operational variables o_t, i_t, o_t^s and i_t^s, respectively. Next, we'll translate these operational variables into operational costs.

Following Holt et al. (1960)'s classic linear-quadratic inventory model[4] (see West, 1985; Eichenbaum, 1989; Blinder and Maccini, 1991; Ramey and West, 1999; Aviv, 2007), we give the manufacturer quadratic inventory and production costs $c_t^i = \gamma_i(i_t - \mu_i)^2$ and $c_t^o = \gamma_o(o_t - \mu)^2$, and we give the supplier quadratic inventory costs $c_t^s = \gamma_s(i_t^s - \mu_s)^2$, where γ_i, γ_o, and γ_s are nonnegative. In expectation, these inventory and production costs are linear in inventory and production variances: $\mathrm{E}(c_t^i) = \gamma_i \mathrm{Var}(i_t)$, $\mathrm{E}(c_t^o) = \gamma_o \mathrm{Var}(o_t)$, and $\mathrm{E}(c_t^s) = \gamma_s \mathrm{Var}(i_t^s)$. Accordingly, the firms pursue the inventory policies that best stabilize these variables:

$$\text{Manufacturer:} \quad \min_{A} \quad \mathrm{E}(c_t^i + c_t^o + \theta c_t^s) = \min_{A} \quad \gamma_i \mathrm{Var}(i_t) + \gamma_o \mathrm{Var}(o_t) + \theta \gamma_s \mathrm{Var}(i_t^s), \tag{6}$$

$$\text{s.t.} \quad \iota' A = \iota'. \tag{7}$$

$$\text{Supplier :} \quad \min_{A_s} \quad \mathrm{E}(c_t^s) = \min_{A_s} \quad \gamma_s \mathrm{Var}(i_t^s), \tag{8}$$

$$\text{s.t.} \quad \iota' A_s = \iota'. \tag{9}$$

Parameter $\theta \geq 0$, in line (6), denotes the degree of supply chain integration, the manufacturer's regard for the supplier's inventory costs. Such upstream inventory consideration could manifest from a cost sharing contract (Cachon, 2003), a reputation penalty associated with production forecast revisions (Cohen et al., 2003; Terwiesch et al., 2005), or the threat of increased wholesale prices or stockout rates (Christopher and Peck, 2003; Lee, 2004).

The next section derives the unique Nash equilibrium of system (6)-(9).

2.4 Nash Equilibrium and Empirical Specification

Taking (8)'s first order conditions (see Appendix B), we find the supplier chooses $A_s = (e_0 e_0' + I - C_0^{-1\prime})^{\psi}$, regardless of what the manufacturer chooses. This IRF matrix corresponds to the classic base stock inventory policy whereby the supplier keeps its ψ-period-ahead inventory forecasts fixed. Plugging this IRF matrix into (5) yields $i_t^s = u_t = \mu_s + \sum_{l=0}^{H+\psi-1} e_l' C_\psi (D_\psi I_\psi (e_0 e_0' + I - C_0^{-1\prime})^{\psi} - I_\psi)\epsilon_{t-l}^o$, where u_t is the cumulative surprise in orders across the supplier's procurement horizon. We use the variance of u_t—the supplier's "exposure period uncertainty" (Chen and Lee, 2009, 2012, p. 786, p. 778)—as our PU measure. Thus, we measure PU with the implied variance of supplier inventories, given that the supplier follows a classic base stock inventory policy, with lead time ψ.[5] The supplier's costs grow linearly with PU.

Henceforth, however, we disregard the supplier, which we introduced simply to motivate our PU measure. Indeed, rather than the supplier's lead time, ψ now parameterizes the manufacturer's PU measure, determining which production signals are *surprising*. A production signal is surprising if and only if its information lead time—the length of time between the realization of the production signal and the realization of the corresponding production (Bray and Mendelson, 2012)—is shorter than ψ. While all production signals contribute to PV, only surprising production signals contribute to PU.[6]

[4] Alternatively, we could specify newsvendor inventory and production costs, in the vein of Balakrishnan et al. (2004). Under this specification, (10)'s objective is linear sum of standard deviations, rather than a linear sum of variances, and (11)'s solution is a fixed point, rather than an explicit expression. Estimating this alternative specification, however, requires implementing a complex nested fixed point estimator. So, for ease of use, we present the simpler linear-quadratic specification.

[5] Note, PU can exceed PV when $\psi > 1$. For example, when the manufacturer's production is an unforecastable *i.i.d.* random variable, then PU is exactly ψ times PV, as each manufacturer production quantity affects ψ supplier inventory levels.

[6] Amongst surprising signals, those with shorter information lead times contribute to PU more than those with longer

Plugging $i_t^s = u_t$ into (6) removes the supplier from the manufacturer's objective, which becomes:

$$\begin{aligned}
\min_A \quad & IV(A) + \alpha PV(A) + \beta PU(A), && (10)\\
\text{s.t.} \quad & \iota' A = \iota', \\
& \alpha = \gamma_o/\gamma_i, \quad \beta = \theta\gamma_s/\gamma_i, \\
& PV(A) = \text{Var}(o_t) = \text{Tr}[A\Sigma A' + A\Gamma + \Gamma' A' + \Lambda], \\
& PU(A) = \text{Var}(u_t) = \text{Tr}[C_\psi(D_\psi I_\psi(e_0 e_0' + I - C_0^{-1\prime})^\psi - I_\psi)(A\Sigma A' + A\Gamma + \Gamma' A' + \Lambda) \\
& \qquad (D_\psi I_\psi(e_0 e_0' + I - C_0^{-1\prime})^\psi - I_\psi)' C_\psi'], \\
\text{and} \quad & IV(A) = \text{Var}(i_t) = \text{Tr}[C_\phi(D_\phi I_\phi A - I_\phi)\Sigma(D_\phi I_\phi A - I_\phi)' C_\phi' + C_\phi(D_\phi I_\phi A - I_\phi)\Gamma I_\phi' D_\phi' C_\phi' \\
& \qquad + C_\phi D_\phi I_\phi \Gamma'(D_\phi I_\phi A - I_\phi)' C_\phi' + C_\phi D_\phi I_\phi \Lambda I_\phi' D_\phi' C_\phi'].
\end{aligned}$$

The manufacturer is averse to inventory variability (IV), production variability (PV), and production uncertainty (PU).[7] We distill low-level cost parameters γ_i, γ_o, γ_s, and θ—which don't distinctly influence behavior, and are thus empirically unidentified—into high-level preference parameters α and β. Parameter $\alpha \geq 0$ denotes the manufacturer's aversion to PV relative to IV, and parameter $\beta \geq 0$ denotes the manufacturer's aversion to PU relative to IV. System (10) underpins our empirical specification.

Taking (10)'s first order conditions (see Appendix B), we find the manufacturer chooses $A = A^*(\alpha, \beta, \phi, \psi, \Sigma, \Gamma)$ where:

$$\begin{aligned}
A^*(\alpha, \beta, \phi, \psi, \Sigma, \Gamma) &= J + K(M(\alpha, \beta, \phi, \psi)K)^{-1}(K' I_\phi' D_\phi' C_\phi' C_\phi I_\phi - M(\alpha, \beta, \phi, \psi)(J + \Gamma' \Sigma^{-1})), \quad (11)\\
J &= e_{H-1}\iota', \\
K &= (I - J) I_{-1}', \\
L(\psi) &= C_\psi(D_\psi I_\psi(e_0 e_0' + I - C_0^{-1\prime})^\psi - I_\psi), \\
\text{and} \quad M(\alpha, \beta, \phi, \psi) &= K' I_\phi' D_\phi' C_\phi' C_\phi D_\phi I_\phi + \alpha K' + \beta K' L(\psi)' L(\psi).
\end{aligned}$$

Exhibit 2 depicts (11)'s optimal solution. First, when $\alpha = \beta = \phi = 0$, the firm acts like a cross-dock facility, holding no stock, and thus incurring no inventory costs; it does so by setting $A = I$. In this case, a last-minute demand shock translates into a last-minute production shock of equal size. When $\alpha > 0$, the firm smooths PV by dispersing mass throughout A's columns, which spreads the risk of a given demand signal across the production horizon. In this case, a last-minute demand shock influences the entire production schedule, with a gradual decay. We call this PV smoothing strategy *signal mixing*, since production signals comprise a mix of demand signals. When $\beta > 0$, the firm smooths PU by diverting mass away from A's first ψ rows, which attenuates the surprising production signals. In this case, a last-minute demand shock most significantly influences the production quantities ψ periods hence, when schedule changes don't increase PU. We call this PU smoothing strategy *signal delaying*, because production quantities respond to demand signals with a delay, to increase supplier reaction times. The following section develops an algorithm to reverse-engineer α and β from the measured degree of signal mixing and signal delaying.

---

information lead times.

[7] We derive this cost trinity from Graves et al. (1998), Balakrishnan et al. (2004), Aviv (2007), and Chen and Lee (2009).

3 Estimation Procedure

Here we develop estimators of our model primitives, α, β, ϕ, ψ, Σ, Λ, and Γ. We employ a new identification strategy: estimating operational features from DSP signatures. §3.1 presents the conditions under which our model is empirically identified, and §3.2 presents our estimators.

3.1 Identification

For empirical identification, we impose additional structure: First, we suppose $E_t[d_t, \cdots, d_{t+H-1}]' = \Xi x_t + \Psi y_t$ and $E_t[o_t, \cdots, o_{t+H-1}]' = \Xi^o x_t + \Psi^o y_t$, where (i) $E_t[d_t, \cdots, d_{t+H-1}]'$ and $E_t[o_t, \cdots, o_{t+H-1}]'$ are the firm's period t demand and production forecasts, (ii) x_t and y_t are arbitrary-lengthed forecast variable vectors, and (ii) Ξ, Ξ^o, Ψ, and Ψ^o are fixed matrices. To satisfy Graves et al. (1998, p. 37)'s market clearing constraint, we impose $\iota'\Xi = \iota'\Xi^o$ and $\iota'\Psi = \iota'\Psi^o$. Next, we let the forecast variables follow vector autoregressive processes: $x_t = \Theta^x_x x_{t-1} + \Theta^x_y y_{t-1} + \xi_t$ and $y_t = \Theta^y_x x_{t-1} + \Theta^y_y y_{t-1} + \Theta^y_\xi \xi_t + \nu_t$, where (i) Θ^x_x, Θ^x_y, Θ^y_x, Θ^y_y, and Θ^y_ξ are fixed matrices, and (ii) ξ_t and ν_t are orthogonal, mean-zero, *i.i.d.* innovation vectors. Note, y_t may depend on x_t's innovations, but not vice versa. Now, by differencing the forecasts (see Appendix A), we find $\epsilon_t = \Phi\xi_t + \Psi\nu_t$ and $\epsilon^o_t = \Phi^o\xi_t + \Psi^o\nu_t$, where $\Phi = \Xi + \Psi\Theta^y_\xi$ and $\Phi^o = \Xi^o + \Psi^o\Theta^y_\xi$. Finally, we impose our identification conditions:

$$\text{Rank}(\Phi E[\xi_t \xi_t']) = H, \tag{12}$$

$$\text{and} \quad \Phi^o = A\Phi. \tag{13}$$

Conditions (12) and (13) are classic instrumental variables inclusion and exclusion restrictions, stating that x_t's innovations can instrument for demand signals. Condition (12) ensures ξ_t has enough linearly independent variables to describe demand signal vector ϵ_t, and condition (13) ensures ξ_t is orthogonal to noise signal vector η_t. More specifically, (12) and (13) imply that A is the only root of matrix function $f(X) = E[(\epsilon^o_t - X\epsilon_t)\xi_t']$. A classic set of GMM moment conditions, $f(A) = 0$ asserts orthogonality between instrument vector, ξ_t, and error vector, η_t; these moment conditions underpin our estimation procedure.

If (12) and (13) hold, then data $\{d_t, o_t, x_t, y_t | t \in \{0, \cdots, T\}\}$ empirically identify our model. First, these variables trivially identify ϵ_t, ϵ^o_t, and ξ_t (see §3.2). Second, with ϵ_t, ϵ^o_t, and ξ_t, we can approximate f with $\widehat{f}(X) = T^{-1}\sum_{t=0}^{T}(\epsilon^o_t - X\epsilon_t)\xi_t'$, the root of which yields a consistent estimate of A. Third, from ϵ_t, ϵ^o_t, and A, we can calculate η_t, and hence Λ and Γ. Fourth, the correlation between demands and lagged production identifies ϕ, since the manufacturer seeks to match $o_{t-\phi}$ with d_t. Fifth, the dispersion of A's columns—which determines the degree of signal mixing—identifies α. And finally, the mass of A's first few rows—which determines the degree of signal delaying—identifies β and ψ (we glean β from the proportion of demand signals that gets delayed, and we glean ψ from the length of the delay). Note, ψ is moot when $\beta = 0$, so we set $\psi = 0$ when $\beta = 0$, without loss of generality.

3.2 Estimators

Sequentially estimating our model primitives (Newey, 1984), the following algorithm unravels the firm's parameters by reverse-engineering its DSP structure:

1. Estimate forecast matrices Ξ, Ψ, Ξ^o, and Ψ^o by jointly regressing d_t and o_t on the elements of $\{x_t, \cdots, x_{t-H+1}, y_t, \cdots, y_{t-H+1}\}$, under market clearing constraints $\iota'\widehat{\Xi} = \iota'\widehat{\Xi}^o$ and $\iota'\widehat{\Psi} = \iota'\widehat{\Psi}^o$ (see Henningsen and Hamann, 2007).
2. Define forecast estimates $\widehat{E}_t[d_t, \cdots, d_{t+H-1}] = \widehat{\Xi}x_t + \widehat{\Psi}y_t$ and $\widehat{E}_t[o_t, \cdots, o_{t+H-1}] = \widehat{\Xi}^o x_t + \widehat{\Psi}^o y_t$.
3. Define signal estimates $\widehat{\epsilon}_t = \widehat{E}_t[d_t, \cdots, d_{t+H-1}] - \widehat{E}_{t-1}[d_t, \cdots, d_{t+H-1}]$ and $\widehat{\epsilon}^o_t = \widehat{E}_t[o_t, \cdots, o_{t+H-1}] - \widehat{E}_{t-1}[o_t, \cdots, o_{t+H-1}]$.
4. Define innovation estimates $\widehat{\xi}$ as the residuals of regressions of the elements of x_t on the elements of x_{t-1} and y_{t-1}.
5. Define matrices $E = [\widehat{\epsilon}_0, \cdots, \widehat{\epsilon}_T]'$, $E^o = [\widehat{\epsilon}^o_0, \cdots, \widehat{\epsilon}^o_T]'$, and $Z = [\widehat{\xi}_0, \cdots, \widehat{\xi}_T]'$.
6. Estimate the IRF matrix by applying GMM to the $f(A) = 0$ moment conditions:[8]

$$\widehat{A} = E^{o\prime} Z Z' E (E' Z Z' E)^{-1}. \tag{14}$$

7. Estimate the signal covariance matrices with the method of moments, evaluated under $\widehat{A}$:

$$\widehat{\Sigma} = E'E/T, \qquad \widehat{\Lambda} = (E^o - E\widehat{A}')'(E^o - E\widehat{A}')/T, \qquad \text{and} \qquad \widehat{\Gamma} = E'(E^o - E\widehat{A}')/T. \tag{15}$$

8. Estimate preference parameters and lead times by applying GMM to the $f(A) = 0$ moment conditions once again. This time, however, use (11)'s optimal IRF matrix, evaluated under $\widehat{\Sigma}$ and $\widehat{\Gamma}$:

$$\{\widehat{\alpha}, \widehat{\beta}, \widehat{\phi}, \widehat{\psi}\} = \underset{\alpha,\beta,\phi,\psi}{\arg\min} \operatorname{Tr}[(E^{o\prime} - A^*(\alpha, \beta, \phi, \psi, \widehat{\Sigma}, \widehat{\Gamma})E')ZZ'(EA^*(\alpha, \beta, \phi, \psi, \widehat{\Sigma}, \widehat{\Gamma})' - E^o)]. \tag{16}$$

9. Reestimate the IRF matrix by plugging the primitive estimates into (11):

$$\widehat{A}^* = A^*(\widehat{\alpha}, \widehat{\beta}, \widehat{\phi}, \widehat{\psi}, \widehat{\Sigma}, \widehat{\Gamma}). \tag{17}$$

Our procedure yields two IRF matrix estimators: $\widehat{A}$ and $\widehat{A}^*$. While both estimators satisfy the $\iota'A = \iota'$ market clearing constraint, only the latter satisfies (11). This overidentifying restriction makes our model falsifiable: $\widehat{A}$ will converge to $\widehat{A}^*$ if and only if our theory is correct. Assuredly, we can estimate the four matrices that govern DSP—Σ, Λ, Γ, and A—without imposing (11).[9] Appendix C illustrates the consistency of our estimators with a Monte Carlo simulation.

[8] Expressions (14) and (16) suppose that we set the GMM weighting matrix to the identity matrix. With general GMM weighting matrix W, these estimators become $\widehat{A} = \arg\min_A \operatorname{vec}[(E^{o\prime} - AE')Z]'W\operatorname{vec}[(E^{o\prime} - AE')Z]$ and $\{\widehat{\alpha}, \widehat{\beta}, \widehat{\phi}, \widehat{\psi}\} = \arg\min_{\alpha,\beta,\phi,\psi} \operatorname{vec}[(E^{o\prime} - A^*(\alpha, \beta, \phi, \psi, \widehat{\Sigma}, \widehat{\Gamma})E')Z]'W\operatorname{vec}[(E^{o\prime} - A^*(\alpha, \beta, \phi, \psi, \widehat{\Sigma}, \widehat{\Gamma})E')Z]$, where $\operatorname{vec}[\cdot]$ is matrix vectorization. With $H^4 = 1,296$ elements, however, the optimal GMM weighting matrix is too large to estimate.

[9] Keane (2010, p. 17) explains that:

> [I]f it can be shown that certain key parameters of a structural model are identified even if the functional form assumptions required for FIML estimation are relaxed, then we gain confidence that the model is a useful tool for organizing the world and making predictions.

4 Data

We estimate our empirical model with a sample of auto manufacturing data. §4.1 briefly recaps the empirical auto OM literature. §4.2 describes our auto dataset, and constructs our forecast variables. And §4.3 discusses how these forecast variables satisfy our empirical identification conditions.

4.1 Review of Empirical Automotive OM Literature

Empirical OM researchers have studied the auto industry for decades. Clark and Fujimoto (1989) study auto product development. Cusumano and Takeishi (1991) study the difference between American and Japanese supplier relations, and Helper and Sako (1995) outline the benefits of committing to specific suppliers. Bresnahan and Ramey (1994) estimate auto production cost functions. Huson and Nanda (1995) correlate Just In Time manufacturing practices with financial performance. Hall (2000) show that kinks in production cost functions lead to production scheduling heterogeneity. Goyal et al. (2006) document a herding effect in auto production flexibility expenditures. Ramey and Vine (2006) study how sales persistence drives production volatility in the auto industry. Cachon and Olivares (2010) attribute three quarters of the difference between the inventories of GM and Toyota to GM's sprawling distribution network. Moreno and Terwiesch (2011) find that firms that build flexible plants can afford to offer fewer price discounts. Guajardo et al. (2012) estimate that increasing a car's price by 1% has a commensurate effect on demand as increasing its warranty length by 9%. Cachon et al. (2012a,b) show that inclement weather slows auto production, and that increasing the number of cars at a dealership increases sales if and only if doing so expands the set of available models. And Gopal et al. (2013) estimate that introducing a new car model decreases plant productivity by 12%-15%.

4.2 Sample Construction

We use automotive data from the Wards Auto InfoBank, maintained by the WardsAuto Group. The data provide sales and period-end inventory levels of cars produced in the U.S. for domestic consumption, measured in physical units, at the model level. The data span from 1985 to 2012, with monthly frequency.[10] We proxy for demand with sales,[11] and derive production with the law of material conservation: $o_t = d_t + i_t - i_{t-1}$. We discard data with nonpositive sales or production, and remove vehicles that don't have at least ten years of uninterrupted data. Exhibit 3 reports summary statistics of our resulting sample.

Next, we derive the variables necessary to estimate our model (which we'll estimate separately for each car in our dataset). First, we define $D_t = [d_t, d_t^2, \bar{d}_{t,f}, \bar{d}_{t,f}^2, \underline{d}_{t,f}, \bar{d}_{t,i}, \bar{d}_{t,i}^2, \underline{d}_{t,i}]'$, where (i) d_t is the demand of the given car model, (ii) $\bar{d}_{t,f}$ and $\underline{d}_{t,f}$ are the corresponding firm's across-model mean demand and standard deviation demand, respectively, and (iii) $\bar{d}_{t,i}$ and $\underline{d}_{t,i}$ are the auto industry's across-model mean demand and standard deviation demand, respectively. Second, we define vector O_t as the production equivalent of vector D_t. Third, we define $S_t = [t, t^2, t^3, \mathbb{1}_{\text{Jan.}}, \cdots, \mathbb{1}_{\text{Dec.}}]'$ as a vector housing the first three powers of t (to account for trends), and monthly dummies (to account for seasonality). And finally, we define $x_t = D_t$,

[10] Hall (2000, p. 684) explains that "most production decisions for automobile assembly plants are made at the monthly frequency."

[11] Sales closely matches demand in our setting, since auto stockouts are rare: Cachon et al. (2012b, p. 11) estimate a 1% rate of stockout for a given model, at a given dealership, in a given and week.

and $y_t = [O_t', O_{t-1}', D_{t-1}', S_t']'$ (see §3.1). Thus, we forecast demand and production with 47 variables, 16 of which resolve in period t (D_t and O_t), and 31 of which resolve prior to period t (D_{t-1}, O_{t-1}, and S_t).

4.3 Satisfying Identification Conditions

Setting x_t to D_t equivalently sets ξ_t to $D_t - E_{t-1}[D_t]$. Thus, D_t's innovations must satisfy identification conditions (12) and (13). Condition (12) should hold, for reasonable H, because D_t's different variables influence demand forecasts differently. For example, d_t's innovations tend to have relatively more sway over ϵ_t's short-information-lead-timed signals, since product-level fluctuations tend to be more fleeting, and $d_{t,i}$'s innovations tend to have relatively more sway over ϵ_t's long-information-lead-timed signals, since industry-level fluctuations tend to be more sustained. Since x_t has eight elements, (12) can only hold for $H \leq 8$; we set $H = 6$, to give ourselves two extra degrees of freedom.

Next, to satisfy condition (13), our demand variable innovations may only communicate with production signals through demand signals. That is, ϵ_t must mediate all correlations between ξ_t and ϵ_t^o. In practice, this means:

1. ξ_t may not influence ϵ_t^o, other than through ϵ_t. For example, a demand shock influencing production schedules via the price of aluminum would violate (13).
2. ϵ_t^o may not influence ξ_t. For example, a sudden factory fire depressing current demand would violate (13).
3. A confounding variable may not influence both ξ_t and ϵ_t^o. For example, a surprise snowstorm disrupting production and current demand would violate (13).

Note, supposing innovations $D_t - E_{t-1}[D_t]$ satisfy (13) is a weaker assumption than supposing variables D_t satisfy (13). Indeed, our instruments comprise only a modest slice of the total D_t variation—the part that resolves in period t; the rest of the D_t variation—the part that resolves before period t—need not satisfy (13). For example, a period-t factory fire may influence D_{t+1} without violating condition (13). In fact, any production shock may influence demand, as long as it does so with a time lag. Fortunately, most production shocks take at least a month to affect demand, since auto dealerships generally hold several months of inventory.

5 Parameter Estimates

With §3's procedure, we estimate decision variable A and primitives α, β, ϕ, ψ, Σ, Λ, and Γ. Since we use sequential estimators, we calculate standard errors with the block bootstrap (Berkowitz and Kilian, 2000; Hardle et al., 2003). We estimate each car model separately.

5.1 Preference Parameters

Exhibit 4 presents our preference parameter estimates: $\widehat{\alpha}$ and $\widehat{\beta}$. Recall that α denotes the cost of PV relative to IV, and β denotes the cost of PU relative to IV.

The $\widehat{\alpha}$ estimates are largely positive, which indicates an intrinsic aversion to PV, irrespective of PU. Indeed, we estimate that the median car's manufacturer would exchange one unit of PV for 2.29 units of

IV, and the average car's manufacturer would exchange one unit of PV for 3.73 units of IV (the variance of deseasonalized demand is our IV, PV, and PU unit of measure). PV is costly in auto manufacturing because "assembly lines are designed to operate at a particular line rate (vehicles per hour)" (Cachon and Olivares, 2010, p. 204). Accordingly, "[p]lant managers rarely change the number of shifts or the line speed" (Hall, 2000, p. 683), opting instead to "operate the plant at its minimum efficient scale (MES), the rate of production that minimizes average cost" (Copeland and Hall, 2005, p. 233).

The $\widehat{\beta}$ estimates are also largely positive, which indicates an intrinsic aversion to PU, irrespective of PV. Indeed, we estimate that the median car's manufacturer of would exchange one unit of PU for .91 units of IV, and the average car's manufacturer would exchange one unit of PU for 2.18 units of IV. Note, β denotes the *incremental* cost of uncertainty, so auto manufacturers generally consider surprising production revisions, which contribute to both PV and PU, more costly than unsurprising signals, which only contribute to PV. This apparent PU aversion could arise from three sources: (i) manufacturing costs associated with short-notice production revisions, (ii) supplier stockouts, or (iii) supply chain coordination. Our data are silent as to the cause, but the literature points to coordination, as we explain below.

First, the positive $\widehat{\beta}$ could stem from intrinsic manufacturing costs tied to last-minute production schedule changes. For example, PU would be costly if manufacturers had to reserve labor ahead of time. However, such production revision costs seem unlikely, since: (i) Bresnahan and Ramey (1994, p. 600) fail to include any production revision penalties in their exhaustive list of auto production "costs directly spelled out in government regulations, labor contracts, or directly implied by the production technology." (ii) Hall (2000, p. 685) explains that firms can change staffing levels with short notice, since "standard labor contracts are written with a one-week time period in mind." (iii) Copeland and Hall (2005, p. 264) suggest that "there are no adjustment costs to varying the work week of capital over time." And (iv) Ramey and Vine (2006, p. 1880) state that "[w]hile overtime hours and inventory adjustments do affect marginal costs, their use incurs no adjustment costs."

Second, the positive $\widehat{\beta}$ could imply frequent supplier stockouts, since manufacturers can't produce without inputs. However, supplier stockouts appear too rare to significantly influence our estimates: Copeland and Hall (2005, p. 241) report that supply disruptions shut down auto plants for only three days a year, on average.

Finally, the positive $\widehat{\beta}$ could imply supply chain coordination. For example, in our model, the manufacturer's aversion to surprising production fluctuations (i.e., its positive β) stems from its consideration of upstream inventory costs (i.e., its positive θ). Indeed, PU aversion implies upstream inventory consideration in most supply chain coordination models (Cachon, 2003; Chen, 2003).

5.2 Lead Times

Exhibit 5 tabulates the joint and marginal distributions of $\widehat{\phi}$ and $\widehat{\psi}$. Recall that ϕ is the manufacturer's production lead time, and ψ is the minimum information lead time necessary for a production signal to be deemed unsurprising—i.e., for it not contribute to PU. We allow $\widehat{\phi}$ to span between zero and three months, and we allow $\widehat{\psi}$ to span between one and three months (and thus deem all signals with less than a month's notice surprising). However, we set $\widehat{\psi} = 0$ when $\widehat{\beta} < .01$, because ψ is moot when $\beta = 0$.

First, our data take advantage of our model's lead time flexibility: rather than consolidate at a boundary condition, $\widehat{\phi}$ and $\widehat{\psi}$ populate the full range of potential lead time values. Second, we find the $\widehat{\psi}$ estimates generally exceed the $\widehat{\phi}$ estimates. This makes sense, as it should take longer for inventory to flow through a complex supplier network than through an auto assembly plant.

5.3 DSP Matrices

Finally, Exhibit 6 present our DSP matrix estimates, $\widehat{\Sigma}$, $\widehat{\Lambda}$, $\widehat{\Gamma}$, $\widehat{A}$, and $\widehat{A}^*$:

- The $\widehat{\Sigma}$ estimates show that demand signals are positively correlated, and that demand signal informativeness—variability—decreases with information lead time.
- The $\widehat{\Lambda}$ estimates demonstrate that the n_t errors affect production schedules: production depends on more than just demand.
- The $\widehat{\Gamma}$ estimates show that the demand and noise signals are indeed correlated. (The first row of $\widehat{\Gamma}$ is mechanically zero, because $\widehat{\xi}_t$ includes the first element of $\widehat{\epsilon}_t$.)
- The similarity between $\widehat{A}$ and $\widehat{A}^*$ suggests (11)'s solution reasonably describes firm behavior. And the flatness of $\widehat{A}$'s and $\widehat{A}^*$'s columns suggests a substantial degree of production smoothing.

6 Production Smoothing

Now we use our primitive estimates to study production smoothing. In §6.1, we disentangle the phenomenon from the bullwhip effect. In §6.2, we define a new, information-based class of production smoothing. And in §6.3, we present our empirical results.

6.1 Detecting Production Smoothing Amidst the Bullwhip Effect

A firm smooths production if it uses inventories to insulate production schedules from demand fluctuations. Microeconomists predicted such smoothing behavior, since "[i]f marginal production costs are increasing and sales vary over time, a cost-minimizing strategy that equates marginal costs across time periods will smooth production relative to sales" (Blinder and Maccini, 1991, p. 78). Accordingly, researchers have generally tested for production smoothing by measuring whether firms produce less variably than they sell. These tests have failed to capture production smoothing, however, since the bullwhip effect generally makes firms amplify, rather than attenuate, demand volatility.

The bullwhip effect has several drivers (Lee et al., 1997). For example, Kahn (1987) and Lee et al. (1997) show that all it takes for production variability (PV) to exceed demand variability (DV) are positively correlated demand signals, and positive production lead times. Under these weak conditions, the firm's current production quantity responds to all the demand signals that realize within its production horizon; pooling the correlated demand signals, in this manner, inflates PV. Lee et al. (1997, p. 549) call this bullwhip cause "demand signal processing" (DSP). This DSP effect can be significant. For example, suppose $\alpha = 0$, $\beta = 0$, $\phi = 2$, $\psi = 2$, $\Lambda = 0$, $\Gamma = 0$, and $\Sigma = vv'/(v'v)$, where $v = [.9^0, .9^1, \cdots, .9^6]$. This case corresponds to Lee et al. (1997, p. 549)'s specification: The demands nearly follow an AR(1) process, and the η_t noise signals are zero (including non-zero η_t only strengthens our point). Under these parameters, the firm's DV is $\text{Tr}[vv'/(v'v)] = 1$, by design, but its PV is $PV(A^*(0, 0, 2, 2, vv'/(v'v), 0)) = 2.20$. Moreover, increasing α

to 10 only drops PV to $PV(A^*(10,0,2,2,vv'/(v'v),0)) = 1.05$, which still exceeds DV. Thus, the bullwhip effect can amplify variability even when PV is ten times more costly than IV.

Now, with $\alpha = 10$, does the firm smooth production? Not according to the traditional criteria—DV exceeding PV. But this a paradox: Why doesn't the firm smooth production when production stability is its primary concern? The answer is that the bullwhip effect makes benchmarking PV to DV an apples-to-oranges comparison. A more accurate PV smoothing measure is the difference between what PV *would be* if the firm had no incentive to smooth, and what PV *actually is*. With this measure, we find the firm actively smooths PV by $PV(A^*(0,0,2,2,vv'/(v'v),0)) - PV(A^*(10,0,2,2,vv'/(v'v),0)) = 2.20 - 1.05 = 1.15$ units (note, we always use DV as our production smoothing unit of measure). Accordingly, our empirical PV smoothing metric is $\Delta\widehat{PV} = PV(A^*(0,0,\widehat{\phi},\widehat{\psi},\widehat{\Sigma},\widehat{\Gamma})) - PV(A^*(\widehat{\alpha},\widehat{\beta},\widehat{\phi},\widehat{\psi},\widehat{\Sigma},\widehat{\Gamma}))$, where $PV(A^*(0,0,\widehat{\phi},\widehat{\psi},\widehat{\Sigma},\widehat{\Gamma}))$ is what PV theoretically would be if the firm were indifferent to production stability, and $PV(A^*(\widehat{\alpha},\widehat{\beta},\widehat{\phi},\widehat{\psi},\widehat{\Sigma},\widehat{\Gamma}))$ is what PV actually is.

This new smoothing metric resolves the false dichotomy between production smoothing and the bullwhip effect. Until now, these distinct phenomena shared a common measure—the difference between PV and DV—which forced them to be antithetical to one another. With our new smoothing measure, we no longer must pit production smoothing against the bullwhip effect—we can embrace both.[12]

6.2 Smoothing Uncertainty

The production smoothing literature has only ever considered PV dampening. Yet firms might smooth production to dampen PU, rather than PV. So we create an equivalent PU smoothing measure: $\Delta\widehat{PU} = PU(A^*(0,0,\widehat{\phi},\widehat{\psi},\widehat{\Sigma},\widehat{\Gamma})) - PU(A^*(\widehat{\alpha},\widehat{\beta},\widehat{\phi},\widehat{\psi},\widehat{\Sigma},\widehat{\Gamma}))$, where $PU(A^*(0,0,\widehat{\phi},\widehat{\psi},\widehat{\Sigma},\widehat{\Gamma}))$ is what PU theoretically would be if the firm were indifferent to production stability, and $PU(A^*(\widehat{\alpha},\widehat{\beta},\widehat{\phi},\widehat{\psi},\widehat{\Sigma},\widehat{\Gamma}))$ is what PU actually is. This measure captures production smoothing's informational aspect—how much smoothing increases the information lead time of production forecast revisions.

Since there's no free lunch, decreasing PV and PU must increase IV. We measure this effect with an equivalent inventory smoothing metric: $\Delta\widehat{IV} = IV(A^*(0,0,\widehat{\phi},\widehat{\psi},\widehat{\Sigma},\widehat{\Gamma})) - IV(A^*(\widehat{\alpha},\widehat{\beta},\widehat{\phi},\widehat{\psi},\widehat{\Sigma},\widehat{\Gamma}))$, where $IV(A^*(0,0,\widehat{\phi},\widehat{\psi},\widehat{\Sigma},\widehat{\Gamma}))$ is what IV theoretically would be if the firm were indifferent to production stability, and $IV(A^*(\widehat{\alpha},\widehat{\beta},\widehat{\phi},\widehat{\psi},\widehat{\Sigma},\widehat{\Gamma}))$ is what IV actually is. This measure captures the inventory costs smoothing production.

6.3 Results

Exhibit 7 tabulates the mean and median $\Delta\widehat{PV}$, $\Delta\widehat{PU}$, and $\Delta\widehat{IV}$ estimates:

- We find auto manufacturers exhibit both the bullwhip effect—PV exceeds DV in over 98% of our sample—and production smoothing—relative to the no-smoothing-incentive scenario, 94% of our sample smooths PV by at least 5% of DV, and 95% smooths PU by at least 5% of DV.
- Our large $\widehat{\alpha}$ and $\widehat{\beta}$ estimates yield large smoothing estimates: We estimate that the average firm smooths PV by 108% of DV, and smooths PU by 374% of DV. If firms didn't smooth production, the median PV would be 67% higher, and the median PU would be 155% higher.

[12]For instance, the firm in the example above both bullwhips and smooths production, when $\alpha = 10$.

- Stabilizing production destabilizes inventories: If firms didn't smooth production, the median IV would be 32% lower. However, the average auto manufacturer appears to receive a favorable "exchange rate," trading 1.08 units of PV and 3.74 units of PU for only 1.93 units of IV.
- We find both traditional mode of smoothing—tempering PV—and the new, information-based mode—tempering PU.

7 Explanatory Regressions

Before concluding, we explore the variation in $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\phi}$, $\widehat{\psi}$, $\Delta\widehat{PV}$, $\Delta\widehat{PU}$, and $\Delta\widehat{IV}$. We first analyze these estimates' cross-sectional variation, and then we analyze their temporal variation.

To explore cross-sectional variations, we regress our estimates on nine vehicle-level regressors. The first five regressors are an intercept and dummy variables for Asian models, European models, luxury models, and trucks. The sixth is company size, measured by the log of average firm sales. The seventh is a proxy for capacity utilization: the average ratio of the vehicle's monthly production and its annual maximum production (this proxy will be sharp as long as production approaches maximum capacity at least once a year). The eighth regressor is Cachon and Olivares (2010)'s production inflexibility measure: the mean absolute difference between a vehicle's production and sales, as a fraction of its average sales. Larger differences between production and sales implies less production flexibility, since "a plant's production should track sales more closely as it becomes more flexible." And the ninth regressor is the log of the brand's average number of dealerships, which we collect from *Automotive News*. Our last two regressors "explain most of the difference in inventory between Toyota and Chrysler, Ford, and GM" (Cachon and Olivares, 2010, p. 215).

Exhibit 8 tabulates the regression coefficients:

- Relative to American manufactures, Asian and European manufacturers exhibit higher consideration for PV relative to IV.
- Larger firms manufacture more quickly and less variably, which suggests economies of scale in production.
- Production stability increases with utilization levels, as capacity constrained firms operate under narrower production rate bands. And manufacturing lead times likewise increase with utilization levels, as excess inventories lengthen cycle times.
- Production tracks demand less closely—i.e., Cachon and Olivares (2010)'s production inflexibility measure increases—as manufacturing lead times and the degree of production smoothing increase.

To explore temporal variations, we reestimate our model in an early subsample and a late subsample. The early subsample captures the older half of each vehicle time series, and the late subsample captures the younger half. For example, our Honda Civic time series comprises the 338 months between February, 1985 and March, 2013, so we allocate the first 169 months—February, 1985 to January, 1999—to the early sample, and the last 169 months—February, 1999 to March, 2013—to the late subsample. By design, both subsamples have the same number of observations of each vehicle. We estimate our model independently for each vehicle in each subsample. To accommodate the shortened time horizons, we streamline our empirical specification, decreasing H from 6 to 5, removing the eight standard deviation forecast variables, and reducing the minimum observation threshold from 120 months to 100 months. Overall, 44 vehicles satisfy this 100-observations-per-subsample requirement; our subsamples include only these 44 vehicles. The interquartile range of the early subsample dates is February, 1991 to March, 1998, and interquartile range of the late subsample dates is January, 2003 to June, 2009.

Exhibit 9 tabulates our estimates' temporal changes. First, $\widehat{\psi}$ has increased significantly: presumably, manufacturers seek longer information lead times as they coordinate with higher supply chain echelons. Second, the degree of PU smoothing has increased significantly, at the expense of inventory stability.

8 Conclusion

Until now, researchers have tested for production smoothing by measuring whether production is less variable than demand. This traditional measure has two problems: First, the bullwhip effect makes demand variability a nearly unattainable benchmark. For example, production variability exceeds demand variability in over 98% of our raw and deseasonalized samples; does this mean auto manufacturers disregard production stability over 98% of the time? Surely not. Second, the traditional measure confounds smoothing production uncertainty (PU) with smoothing production variability (PV), obscuring smoothing's operational effects. To address these issues, we develop new metrics that measure the amount firms actively stabilize PV and PU, relative to the scenario in which firms have no impetus to smooth production. Robust to the bullwhip effect, these measures indicate firms smooth production substantially, despite production being more volatile than demand. And they indicate firms smooth both PV and PU. Although previously undocumented, we find PU smoothing a substantial phenomenon, more than three times as large as PV smoothing, and increasing significantly over time. This phenomenon suggests production smoothing has an informational aspect.

We estimate that the average auto manufacturer considers PV and PU as 3.7 and 2.2 times as costly as inventory variability, respectively. The PV aversions allude to rigid auto production processes (see Bresnahan and Ramey, 1994; Copeland and Hall, 2005). And the PU aversions allude to supply chain coordination—manufacturers tempering orders to stabilize upstream inventories (see Graves et al., 1998; Balakrishnan et al., 2004; Aviv, 2007; Chen and Lee, 2009). We identify the PV and PU aversions from two artifacts in our data: First, when setting production schedules, auto manufacturers *mix* demand fluctuations—even those they anticipate well ahead of time. This signal mixing implies an inherent aversion to PV: dispersing unsurprising demand signals doesn't decrease PU, but does decrease PV. Second, auto manufacturers *delay* production revisions, keeping the production plan relatively stable in the final few months. This signal delaying implies an incremental aversion to PU: postponing production changes doesn't decrease PV, but does decrease PU.

This signal processing identification strategy is new. Indeed, this work establishes a new empirical approach: gleaning operational insights from DSP signatures. This approach can yield other inventory management insights, as the transformation of demand into production is an intricate process that speaks volumes about operations. To implement this approach, we develop an empirical production scheduling model, based on the Generalized Order-Up-To Policy (Graves et al., 1998; Chen and Lee, 2009). The model treats firms as signal processing machines.

Appendix

A Martingale Model of Forecast Evolution

The lynchpin of our signal processing specification is the Martingale Model of Forecast Evolution (MMFE) of Hausman (1969) and Heath and Jackson (1994). All an MMFE requires is (i) a discrete stochastic process $\{x_t\}_{-\infty}^{\infty}$, in which random variable x_t resolves in period t, (ii) a forecaster that updates its conditional expectations of the unresolved random variables in each period, and (iii) the forecaster's period t information set to subsume its period $t-1$ information set. With these assumptions, we can decompose a random variable into a series of uncorrelated forecast revisions: $x_t = \bar{x}+\sum_{l=0}^{\infty}(E_{t-l}x_t - E_{t-l-1}x_t)$, where $\bar{x}$ is x_t's apriori mean. The forecast revisions, $E_{t-l}x_t - E_{t-l-1}x_t$, are mean-zero and uncorrelated, because the forecasts, $E_{t-l}x_t$, follow a Martingale: $E_s x_t = E_s(E_s[x_t|E_\tau x_t]) = E_s[E_\tau x_t]$, for $s \leq \tau \leq t$. Following Graves et al. (1998, p. S37), we further suppose (i) the forecast revisions are covariance stationary, and (ii) that the forecaster has an H period long forecast horizon. With this, we can express an MMFE as $x_t = \bar{x} + \sum_{l=0}^{H-1} e_l' u_{t-l}$, where $u_t = E_t[x_t, \cdots, x_{t+H-1}]' - E_{t-1}[x_t, \cdots, x_{t+H-1}]'$ has fixed covariance matrix Ω. Note that u_t is uncorrelated with u_τ, for $\tau \neq t$, and that $\mathrm{Var}(x_t) = \mathrm{Tr}(\Omega)$.

B Deriving the Nash Equilibrium

Defining $J = e_{H-1}\iota'$ and $K = (I-J)I_{-1}'$, we can express any A_s that satisfies constraint (9) as $J+K\tilde{A}_s$, for some $(H-1)\times H$ matrix $\tilde{A}_s$. This fact enables us to reformulate the supplier's objective as an unconstrained optimization problem:

$$\min_{\tilde{A}_s} \gamma_s \mathrm{Tr}[C_\psi(D_\psi I_\psi(J+K\tilde{A}_s) - I_\psi)(A\Sigma A' + A\Gamma + \Gamma' A' + \Lambda)(D_\psi I_\psi(J+K\tilde{A}_s) - I_\psi)' C_\psi]. \tag{18}$$

We calculate (18)'s first order conditions with the following identities: $\partial \mathrm{Tr}[AXBX'A']/\partial X = A'AXB + A'AXB'$ and $\partial \mathrm{Tr}[AXB]/\partial X = A'B'$. When $\Sigma^o = A\Sigma A' + A\Gamma + \Gamma' A' + \Lambda$ is invertible, the only matrix that satisfies (18)'s first order conditions is $\tilde{A}_s = (K'I_\psi' D_\psi' C_\psi' C_\psi D_\psi I_\psi K)^{-1} K' I_\psi' D_\psi' C_\psi' C_\psi (I_\psi - D_\psi I_\psi J)$. This must be the optimal solution for all Σ^o, since (18)'s objective is bowl-shaped in $\tilde{A}_s$, and any singular matrix can be approximated arbitrarily closely with an invertible matrix. Thus, regardless of the manufacturer's decision, the supplier chooses $A_s = J + K(K'I_\psi' D_\psi' C_\psi' C_\psi D_\psi I_\psi K)^{-1} K' I_\psi' D_\psi' C_\psi' C_\psi (I_\psi - D_\psi I_\psi J) = (e_0 e_0' + I - C_0^{-1\prime})^\psi$.

Setting $A_s = (e_0e_0' + I - C_0^{-1\prime})^\psi$ and $A = J + K\tilde{A}$ yields a similarly unconstrained manufacturer objective:

$$\min_{\tilde{A}} IV(J+K\tilde{A}) + \alpha PV(J+K\tilde{A}) + \beta PU(J+K\tilde{A}). \tag{19}$$

We calculate (19)'s first order conditions with the two identities mentioned above. The only matrix that satisfies these conditions is $\tilde{A} = \Big(K'I_\phi' D_\phi' C_\phi' C_\phi D_\phi I_\phi K + \alpha K'K + \beta K'(D_\psi I_\psi (e_0e_0' + I - C_0^{-1\prime})^\psi - I_\psi)' C_\psi' C_\psi (D_\psi I_\psi (e_0 e_0' + I - C_0^{-1\prime})^\psi - I_\psi)K\Big)^{-1} \cdot \Big(K'I_\phi' D_\phi' C_\phi' C_\phi - \big(K' I_\phi' D_\phi' C_\phi' C_\phi D_\phi I_\phi + \alpha K' + \beta K'(D_\psi I_\psi (e_0e_0' + I - C_0^{-1\prime})^\psi - I_\psi)' C_\psi' C_\psi (D_\psi I_\psi (e_0e_0' + I - C_0^{-1\prime})^\psi - I_\psi)\big)(J + \Gamma' \Sigma^{-1})\Big)$. This must be the optimal solution, since (19)'s objective is bowl-shaped in $\tilde{A}$.

C Monte Carlo Simulation

We now conduct a brief Monte Carlo simulation study to illustrate the consistency of our estimators. We first independently draw 200 sets of primitives from a set of distributions: We specify α to be uniform across $[0,2]$, β to be uniform across $[0.1,2]$, ϕ to be uniform across $\{0,1,2,3\}$, and ψ to be uniform across $\{1,2,3\}$. And we specify $\Omega = \Phi E[\xi_t \xi_t']$ and $\begin{bmatrix} \Sigma - \Omega & \Gamma \\ \Gamma' & \Lambda \end{bmatrix}$ to follow Wishart distributions, with identity scale matrices, and $H = 6$ and $2H = 12$ degrees of freedom, respectively (Rossi et al., 2005, p. 30). For each set of primitives, we then (i) draw $T \in \{100,\ 500,\ 2500\}$ realizations of $[\epsilon_t', \eta_t', \xi_t']'$ from a mean-zero multivariate normal, with covariance matrix $\begin{bmatrix} \Sigma & \Gamma & \Omega \\ \Gamma' & \Lambda & 0 \\ \Omega' & 0 & \Omega \end{bmatrix}$, (ii) calculate the optimal IRF matrix with (11), (iii) derive the corresponding ϵ_t^o production signals with (2), (iv) construct matrices E, E^o, and Z from vectors ϵ_t, ϵ_t^o, and ξ_t, and (v) estimate the primitives from E, E^o, and Z with (14)-(16).

Exhibit 10 plots α against $\widehat{\alpha}$ and β against $\widehat{\beta}$. Since the variance of the error terms are arbitrary, the magnitude of the estimation errors are as well. What's important, however, is that $\widehat{\alpha}$ and $\widehat{\beta}$ converge to α and β, without bias. The $\widehat{\phi}$ and $\widehat{\psi}$ estimates likewise converge: as T increases from 100 to 500 to 2,500, the fraction of the time $\widehat{\phi}$ equals ϕ increases from 97% to 99% to 100%, and the fraction of the time $\widehat{\psi}$ equals ψ increases from 86% to 96% to 99%. Note, convergence in latter stage estimators $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\phi}$, and $\widehat{\psi}$ implies convergence in former stage estimators $\widehat{\Sigma}$, $\widehat{\Lambda}$, $\widehat{\Gamma}$ and $\widehat{A}$.

Exhibit 1: Glossary of Variables

Parameter	Description	Section
A, A_s	Manufacturer's and supplier's IRF matrices, or signal routing matrices; their decision variables	2.2
$A^*(\cdot)$	Manufacturer's optimal IRF matrix, as a function of model primitives	2.4
α	Manufacturer's aversion to production variability (PV) relative to inventory variability (IV)	2.4
β	Manufacturer's aversion to production uncertainty (PU) relative to inventory variability (IV)	2.4
C_x	A cumulative sum operator; an $(H+x) \times (H+x)$ lower triangular matrix of ones	2.2
c_t^i, c_t^o, c_t^s	Manufacturer's and supplier's period t production and inventory costs	2.3
D_x	An x period delay operator, $(I - C_x^{-1})^x$	2.2
d_t	Manufacturer's period t demand	2.2
e_l	A vector with a one in its $(l+1)^{th}$ position and zeros elsewhere	2.2
ϵ_t	Manufacturer's demand signal vector, housing demand forecast revisions	2.2, 2.3
ϵ_t^o, $\epsilon_t^{o,s}$	Manufacturer's and supplier's production signal vectors, housing production forecast revisions	2.2, 2.3
ϵ_t^i, $\epsilon_t^{i,s}$	Manufacturer's and supplier's inventory signal vectors, housing inventory forecast revisions	2.2, 2.3
η_t	Noise signal vector, housing forecast revisions of error term n_t	2.2, 2.3
Γ	Covariance matrix between demand signal vector ϵ_t and noise signal vector η_t	2.2
γ_i, γ_o, γ_s	Parameters that underlie the manufacturer's and supplier's production and inventory costs	2.3
H	Length of forecast horizon and production scheduling horizon	2.2
i_t, i_t^s	Manufacturer's and supplier's period t inventory levels	2.2
I_x	A matrix reshape operator; an $(H+x) \times H$ version of the identity matrix	2.2
$IV(\cdot)$	Manufacturer's inventory variability (IV), as a function of IRF matrix A	2.4
ι	An H-lengthed vector of ones, $\sum_{l=0}^{H-1} e_l$	2.2
Λ	Covariance matrix of noise signal vector η_t	2.2
μ	Mean demand and production level	2.2
μ_i, μ_s	Manufacturer's and supplier's mean inventory levels	2.2
n_t	Empirical error term that shifts production, and decomposes into η_t noise signals	2.2
o_t, o_t^s	Manufacturer's and supplier's period t production quantities	2.2
ϕ	Manufacturer's lead time	2.1
ψ	Supplier's lead time; minimum information lead time of "unsurprising" production signals	2.1, 2.4
$PU(\cdot)$	Manufacturer's production uncertainty (PU), as a function of IRF matrix A	2.4
$PV(\cdot)$	Manufacturer's production variability (PV), as a function of IRF matrix A	2.4
Σ	Covariance matrix of demand signal vector ϵ_t	2.2
θ	Degree of supply chain coordination; amount the manufacturer considers supplier inventory costs	2.3
u_t	Cumulative surprise in the manufacturer's production across the supplier's procurement horizon	2.4

Exhibit 2: Optimal IRF Matrices and Corresponding Production Schedules

The left panel depicts expression (11)'s solution, $A^*(\alpha, \beta, \phi, \psi, \Sigma, \Gamma)$. The n^{th} curve of each plot traces the n^{th} column of $A^*(\cdot)$, the impulse response function governing how demand signals with $(n-1)$ information lead times transform into production quantities. The right panel depicts how the corresponding production schedules change in response to an impulse demand signal, $\epsilon_t = e_0 = [1, 0, \cdots, 0]'$ (i.e., a last-minute demand shock); specifically, this panel plots the first column of $A^*(\cdot)$. For each panel, we consider four cases, with α and β in $\{0, 4\}$, and $\Sigma = I$, $\Gamma = 0$, $\phi = 0$, and $\psi = 2$. The vertical dotted lines, at $x = \psi + 1$, denote the cutoff between surprising and unsurprising production signals: only mass allocated to the left of these lines contributes to PU. As β increases, $A^*(\cdot)$'s curves shift rightward, which postpones production. We call this PU smoothing technique *signal delaying*. As α increases, $A^*(\cdot)$'s columns flatten, which more evenly distributes work across the production horizon. We call this PV smoothing technique *signal mixing*.

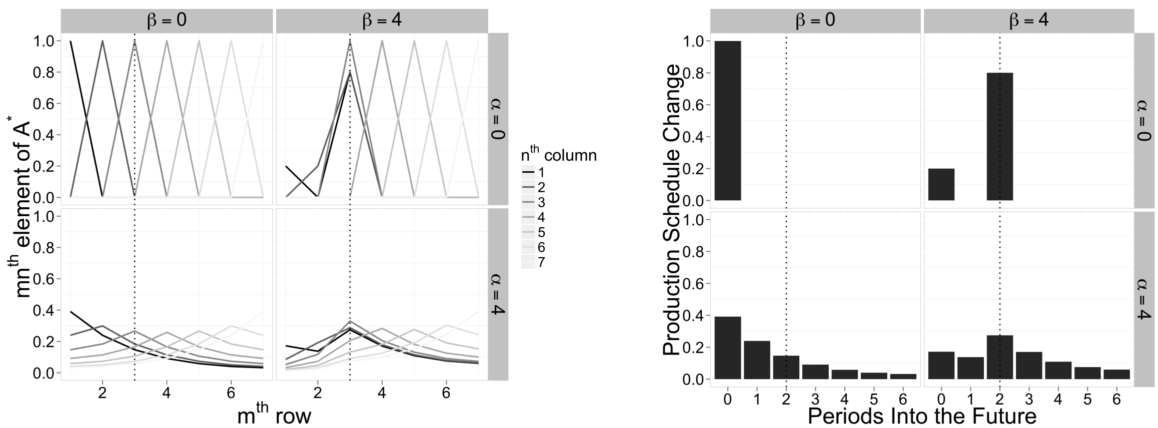


Exhibit 3: Sample Summary Statistics

Our sample only includes car models with at least 120 clean, consecutive observations. This table's sales and inventory figures correspond to monthly, company-level aggregates.

		# Models	# Obs.	Years	Sales Mean	Sales S.D.	Inv. Mean	Inv. S.D.
North America	General Motors	39	6,539	1985-2012	10,253	10,525	28,524	31,117
	Chrysler	16	2,810	1985-2012	11,531	7,691	30,643	24,148
	Ford	15	3,564	1985-2012	17,839	16,001	46,986	46,364
	Total	70	12,913	1985-2012	12,625	12,256	34,081	35,700
Asia	Toyota	19	3,880	1985-2012	9,240	9,259	13,825	14,880
	Nissan	10	1,742	1985-2012	7,556	4,801	20,137	13,180
	Honda	8	1,629	1985-2012	15,450	11,115	26,274	23,845
	Mazda	4	565	1990-2012	4,847	2,116	13,544	4,526
	Mitsubishi	4	605	1995-2012	2,956	2,271	7,970	4,485
	Hyundai	3	511	1997-2012	6,745	4,395	12,865	6,980
	Subaru	3	554	1995-2012	4,483	1,859	9,230	3,673
	Total	51	9,486	1985-2012	8,922	8,643	16,412	16,020
Europe	Volkswagen	4	716	1993-2012	5,263	3,785	16,225	10,838
	BMW	3	623	1992-2012	4,888	2,671	5,917	4,194
	Audi	2	385	1995-2012	2,329	1,199	5,305	3,007
	Daimler	2	282	1996-2008	1,353	733	2,254	1,312
	Porsche	2	379	1991-2011	622	337	915	609
	Total	13	2,385	1991-2012	3,492	3,161	7,685	8,715
Total		134	24,784	1985-2012	10,329	10,762	24,778	29,488

Exhibit 4: Preference Parameter Estimates

We tabulate the mean and median $\widehat{\alpha}$ and $\widehat{\beta}$ estimates. We derive standard errors with the block bootstrap, and asterisk estimates that are significantly greater than zero, with a one-sided $p = 0.01$ test.

		$\widehat{\alpha}$		$\widehat{\beta}$	
		Mean	Median	Mean	Median
North America	Chrysler	1.33	0.94	1.16*	0.64*
		(0.84)	(0.41)	(0.42)	(0.18)
	Ford	4.59*	3.35*	2.43*	1.31*
		(0.89)	(1.22)	(0.48)	(0.25)
	General Motors	4.27*	3.82*	3.95*	2.11*
		(0.37)	(0.41)	(0.53)	(0.31)
	Total	3.67*	2.37*	2.98*	1.38*
		(0.33)	(0.33)	(0.30)	(0.14)
Asia	Honda	6.67*	5.93*	2.88*	1.00*
		(1.22)	(2.54)	(0.89)	(0.40)
	Hyundai	0.86	1.01	0.30	0.35*
		(0.74)	(0.52)	(0.13)	(0.10)
	Mazda	0.31	0.16	0.36	0.19
		(1.05)	(0.40)	(0.30)	(0.20)
	Mitsubishi	0.02	0.00	0.21	0.15
		(0.84)	(0.59)	(0.15)	(0.11)
	Nissan	4.48*	2.56	1.77*	1.42*
		(1.10)	(1.17)	(0.68)	(0.32)
	Subaru	0.01	0.00	0.29	0.34
		(0.70)	(0.05)	(0.76)	(0.15)
	Toyota	5.70*	4.52*	1.49*	0.90*
		(0.93)	(1.21)	(0.27)	(0.21)
	Total	4.13*	2.42*	1.43*	0.78*
		(0.40)	(0.41)	(0.22)	(0.12)
Europe	Audi	3.60	3.60	0.69	0.69
		(2.35)	(2.35)	(1.59)	(1.59)
	BMW	0.84	0.00	0.30	0.00
		(0.78)	(0.12)	(0.21)	(0.09)
	Daimler	0.01	0.01	0.12	0.12
		(0.08)	(0.08)	(0.11)	(0.11)
	Porsche	0.98	0.98	0.52	0.52
		(1.02)	(1.02)	(0.26)	(0.26)
	Volkswagen	5.38*	4.53	1.72	1.51
		(1.99)	(2.81)	(1.29)	(1.09)
	Total	2.56*	0.44	0.80	0.52*
		(0.71)	(0.81)	(0.47)	(0.14)
Total		3.73*	2.29*	2.18*	0.91*
		(0.27)	(0.27)	(0.20)	(0.09)

Exhibit 5: Lead Time Estimates

We tabulate the joint and marginal distributions of $\widehat{\phi}$ and $\widehat{\psi}$. We set $\widehat{\psi} = 0$ if and only if $\widehat{\beta} < .01$ (ψ is moot when $\beta = 0$). We derive standard errors with the block bootstrap.

	$\widehat{\phi} = 0$	$\widehat{\phi} = 1$	$\widehat{\phi} = 2$	$\widehat{\phi} = 3$	All
$\widehat{\psi} = 0$	0.04	0.00	0.01	0.01	0.06
	(0.02)	(0.01)	(0.00)	(0.01)	(0.02)
$\widehat{\psi} = 1$	0.04	0.01	0.02	0.02	0.10
	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)
$\widehat{\psi} = 2$	0.04	0.04	0.03	0.09	0.20
	(0.02)	(0.02)	(0.01)	(0.02)	(0.03)
$\widehat{\psi} = 3$	0.16	0.13	0.09	0.25	0.63
	(0.03)	(0.03)	(0.02)	(0.03)	(0.04)
All	0.29	0.19	0.15	0.37	1.00
	(0.04)	(0.03)	(0.03)	(0.03)	(0.00)

Exhibit 6: DSP Matrix Estimates

These figures plot the quartiles of our DSP matrix estimates. The left panel depicts IRF matrix estimates $\widehat{A}$ and $\widehat{A}^*$, and the right panel depicts signal covariance matrix estimates $\widehat{\Sigma}$, $\widehat{\Lambda}$, and $\widehat{\Gamma}$. We measure signals ϵ_t and η_t in terms of the standard deviation of deseasonalized demand, scaling $\mathrm{Tr}[\widehat{\Sigma}] = 1$. The curves correspond to the matrices' columns.

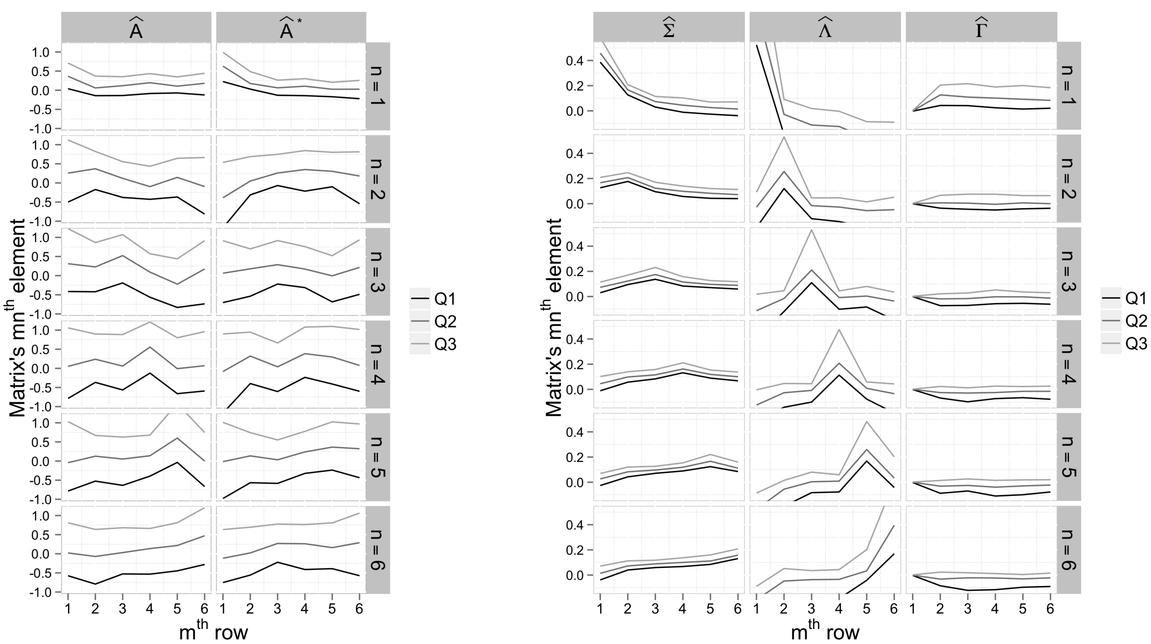


Exhibit 7: Production Smoothing

This table reports the mean and median effects of production smoothing. We express all values as a fraction of deseasonalized demand variability. The top rows correspond to production variability (PV), production uncertainty (PU), and inventory variability (IV) under the counterfactual scenario in which firms disregard production stability: $PV(A^*(0,0,\widehat{\phi},\widehat{\psi},\widehat{\Sigma},\widehat{\Gamma}))$, $PU(A^*(0,0,\widehat{\phi},\widehat{\psi},\widehat{\Sigma},\widehat{\Gamma}))$, and $IV(A^*(0,0,\widehat{\phi},\widehat{\psi},\widehat{\Sigma},\widehat{\Gamma}))$. The middle rows correspond to PV, PU, and IV under the current scenario: $PV(A^*(\widehat{\alpha},\widehat{\beta},\widehat{\phi},\widehat{\psi},\widehat{\Sigma},\widehat{\Gamma}))$, $PU(A^*(\widehat{\alpha},\widehat{\beta},\widehat{\phi},\widehat{\psi},\widehat{\Sigma},\widehat{\Gamma}))$, and $IV(A^*(\widehat{\alpha},\widehat{\beta},\widehat{\phi},\widehat{\psi},\widehat{\Sigma},\widehat{\Gamma}))$. And the bottom rows correspond to the differences: $\Delta\widehat{PV} = PV(A^*(0,0,\widehat{\phi},\widehat{\psi},\widehat{\Sigma},\widehat{\Gamma})) - PV(A^*(\widehat{\alpha},\widehat{\beta},\widehat{\phi},\widehat{\psi},\widehat{\Sigma},\widehat{\Gamma}))$, $\Delta\widehat{PU} = PU(A^*(0,0,\widehat{\phi},\widehat{\psi},\widehat{\Sigma},\widehat{\Gamma})) - PU(A^*(\widehat{\alpha},\widehat{\beta},\widehat{\phi},\widehat{\psi},\widehat{\Sigma},\widehat{\Gamma}))$, and $\Delta\widehat{IV} = IV(A^*(0,0,\widehat{\phi},\widehat{\psi},\widehat{\Sigma},\widehat{\Gamma})) - IV(A^*(\widehat{\alpha},\widehat{\beta},\widehat{\phi},\widehat{\psi},\widehat{\Sigma},\widehat{\Gamma}))$. All figures are significantly different from zero, at the $p = 0.01$ level. We derive standard errors with the block bootstrap.

	PV		PU		IV	
	Mean	Median	Mean	Median	Mean	Median
Counterfactual Scenario	2.92	2.83	6.14	5.79	3.05	2.93
	(0.06)	(0.09)	(0.27)	(0.39)	(0.12)	(0.20)
Current Scenario	1.84	1.69	2.41	2.23	4.98	4.64
	(0.05)	(0.05)	(0.11)	(0.15)	(0.22)	(0.41)
Difference	1.08	1.08	3.74	3.15	−1.93	−1.50
	(0.05)	(0.10)	(0.20)	(0.35)	(0.13)	(0.17)

Exhibit 8: Analysis of Cross-Sectional Variation

We regress $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\phi}$, $\widehat{\psi}$, $\Delta\widehat{PV}$, $\Delta\widehat{PU}$, and $\Delta\widehat{IV}$ on (i) an intercept, (ii) dummy variables for vehicles made by Asian and European companies, luxury cars, and trucks, and (iii) measures for company size, capacity utilization, production inflexibility, and dealership counts. This table provides the regression coefficients, with each row corresponding to a different dependent variable. We derive standard errors with the block bootstrap, and asterisk estimates that are significantly greater than zero, with a two-sided $p = 0.05$ test.

	Intercept	Asian	European	Luxury	Truck	Size	Utilization	Inflexibility	Dealers
$\widehat{\alpha}$	−43.29*	4.10*	6.60*	−0.19	0.01	1.66*	0.27*	0.28*	0.02
	(8.21)	(1.59)	(1.67)	(1.39)	(0.75)	(0.41)	(0.06)	(0.07)	(1.06)
$\widehat{\beta}$	−21.68*	−0.65	−0.25	0.01	0.58	0.32	0.23*	0.30*	−0.41
	(4.73)	(0.71)	(1.14)	(0.72)	(0.40)	(0.28)	(0.05)	(0.05)	(0.45)
$\widehat{\phi}$	−3.00	0.18	−0.39	0.67	0.11	−0.56*	0.06*	0.08*	0.63
	(2.66)	(0.45)	(0.60)	(0.56)	(0.20)	(0.21)	(0.02)	(0.02)	(0.33)
$\widehat{\psi}$	−0.22	0.20	0.94	0.11	0.09	0.15	0.02	−0.01	−0.02
	(1.89)	(0.33)	(0.56)	(0.29)	(0.12)	(0.13)	(0.01)	(0.02)	(0.20)
$\Delta\widehat{PV}$	−3.36*	0.05	−0.17	0.39	−0.03	−0.18	0.05*	0.05*	0.29
	(1.35)	(0.22)	(0.35)	(0.30)	(0.11)	(0.11)	(0.01)	(0.01)	(0.15)
$\Delta\widehat{PU}$	−15.98*	0.67	0.59	2.02	0.52	−0.44	0.18*	0.17*	1.03
	(5.40)	(0.81)	(1.53)	(1.25)	(0.51)	(0.44)	(0.05)	(0.05)	(0.67)
$\Delta\widehat{IV}$	15.45*	−0.60	−0.73	−0.98	−0.45	−0.05	−0.13*	−0.16*	−0.50
	(2.90)	(0.54)	(0.71)	(0.66)	(0.26)	(0.21)	(0.03)	(0.03)	(0.38)

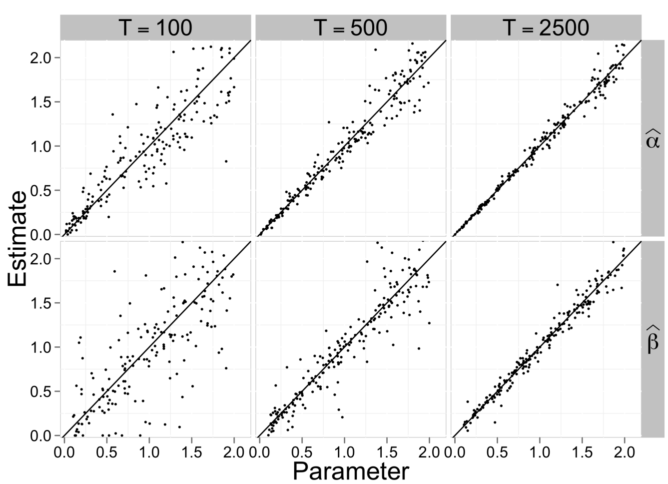
Exhibit 9: Analysis of Temporal Variation

We estimate how estimates $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\phi}$, $\widehat{\psi}$, $\Delta\widehat{PV}$, $\Delta\widehat{PU}$, and $\Delta\widehat{IV}$ change over time by reestimating our model across an early subsample and a late subsample. The early subsample is comprised of the first chronological half of the vehicle time series, and the late subsample is comprised of the second chronological half of the vehicle time series. The subsamples each include 44 vehicles. We tabulate the mean estimates and estimate changes. We derive standard errors with the block bootstrap, and asterisk estimates that are significantly different from zero, with a two-sided $p = 0.05$ test.

	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\phi}$	$\widehat{\psi}$	$\Delta\widehat{PV}$	$\Delta\widehat{PU}$	$\Delta\widehat{IV}$
Early Subsample	2.11*	1.46*	2.07*	2.28*	1.07*	2.99*	−1.29*
	(0.19)	(0.15)	(0.20)	(0.12)	(0.08)	(0.26)	(0.12)
Late Subsample	2.30*	1.69*	1.64*	2.62*	1.06*	3.87*	−1.72*
	(0.20)	(0.15)	(0.13)	(0.10)	(0.09)	(0.32)	(0.14)
Change	0.19	0.23	−0.43	0.34*	−0.01	0.88*	−0.43*
	(0.26)	(0.18)	(0.25)	(0.14)	(0.11)	(0.33)	(0.13)

Exhibit 10: Estimator Simulation Study

We conduct a simulation study of the small-sample performance of our estimators. First, we randomly generate two hundred sets of firm parameters. From a given set of parameters, we then simulate three random datasets, with $T = 100$, 500, and 2,500 observations. Finally, we estimate the underlying parameters from the simulated datasets. The scatterplots below graph the actual simulated preference parameters, α and β, verses their corresponding estimates, $\widehat{\alpha}$ and $\widehat{\beta}$. Deviations from the $y = x$ lines represent estimation error.



References

Aksin, Z, B Ata, S Emadi, CL Su. 2012. Structural estimation of callers' delay sensitivity in call centers. *Management Science* **1909** 1–20. URL http://faculty.chicagobooth.edu/che-lin.su/research/CallCenters_AAES2012.pdf.

Allon, G, A Federgruen, M Pierson. 2011. How much is a reduction of your customers' wait worth? an empirical study of the fast-food drive-thru industry based on structural estimation methods. *Manufacturing & Service Operations ...* **36** 4. URL http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:How+much+is+a+reduction+of+your+customers'+wait+worth?+An+empirical+study+of+the+fast-food+drive-thru+industry+based+on+structural+estimation+methods.#0.

Anand, K. S., H. Mendelson. 1997. Information and organization for horizontal multimarket coordination. *Management Science* **43**(12) 1609–1627.

Aviv, Yossi. 2003. A time-series framework for supply-chain inventory management. *Operations Research* **51**(2) 210–227. doi:10.1287/opre.51.2.210.12780. URL http://www.jstor.org/stable/4132401.

Aviv, Yossi. 2007. On the benefits of collaborative forecasting partnerships between retailers and manufacturers. *Management Science* **53**(5) 777–794. doi:10.1287/mnsc.1060.0654. URL http://mansci.journal.informs.org/cgi/doi/10.1287/mnsc.1060.0654.

Balakrishnan, Anantaram, Joseph Geunes, M.S. Pangburn. 2004. Coordinating supply chains by controlling upstream variability propagation. *Manufacturing & Service Operations Management* **6**(2) 163. doi:10.1287/msom.1030.0031. URL http://msom.highwire.org/cgi/content/abstract/6/2/163.

Beckmann, M. J. 1961. Production smoothing and inventory control. *Operations Research* **9**(4) 456–467. doi:10.1287/opre.9.4.456. URL http://or.journal.informs.org/cgi/doi/10.1287/opre.9.4.456.

Berkowitz, J., L. Kilian. 2000. Recent developments in bootstrapping time series. *Econometric Reviews* **19**(1) 1–48.

Blinder, A. S., L. J. Maccini. 1991. Taking stock: a critical assessment of recent research on inventories. *The journal of economic perspectives* 73–96.

Boute, Robert N, Jan Van Mieghem. 2011. Global Dual Sourcing and Order Smoothing.

Bray, Robert L., Haim Mendelson. 2012. Information Transmission and the Bullwhip Effect : An Empirical Investigation. *Management Science* **58**(5) 860–875.

Bresnahan, TF, VA Ramey. 1994. Output Fluctuations at the Plant Level. *The Quarterly Journal of ...* **109**(3) 593–624. URL http://qje.oxfordjournals.org/content/109/3/593.short.

Cachon, G, Santiago Gallino, Marcelo Olivares. 2012a. Severe weather and automobile assembly productivity. *Available at SSRN 2099798* URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2099798.

Cachon, G P. 2003. Supply chain coordination with contracts. *Handbooks in operations research and management science* **11** 229–340.

Cachon, G. P., T. Randall, G. M. Schmidt. 2007. In search of the bullwhip effect. *Manufacturing & Service Operations Management* **9**(4) 457–479. doi:10.1287/msom.1060.0149. URL http://msom.journal.informs.org/cgi/doi/10.1287/msom.1060.0149.

Cachon, G.P., S. Gallino, M. Olivares. 2012b. Does inventory increase sales? The billboard and scarcity effects in U.S. automobile dealerships.

Cachon, G.P., M. Olivares. 2010. Drivers of finished-goods inventory in the US automobile industry. *Management Science* **56**(1) 202.

Chen, F. 2003. Information sharing and supply chain coordination. *Handbooks in Operations Research and Management Science* **11**(2000) 341–421. URL http://linkinghub.elsevier.com/retrieve/pii/S0927050703110079.

Chen, L., H. L. Lee. 2009. Information sharing and order variability control under a generalized demand model. *Management Science* **55**(5) 781–797. doi:10.1287/mnsc.1080.0983. URL http://mansci.journal.informs.org/cgi/doi/10.1287/mnsc.1080.0983.

Chen, Li, Hau L Lee. 2012. Bullwhip effect measurement and its implications. *Operations Research* **60**(4) 771–784.

Christopher, M., H. Peck. 2003. *Marketing Logistics*. Butterworth-Heinemann.

Clark, Kim B KB, Takahiro Fujimoto. 1989. Lead time in automobile product development explaining the Japanese advantage. *Journal of Engineering and Technology Management* **6**(1) 25–58. doi:10.1016/0923-4748(89)90013-1. URL http://www.sciencedirect.com/science/article/

B6VF3-4593CX4-T/2/d39fbedfa98ecc288f1a7d2df573b294http://www.sciencedirect.com/science/article/pii/0923474889900131.

Cohen, M. A. MA, TH H. Ho, Z. J. Ren, C. Terwiesch. 2003. Measuring imputed cost in the semiconductor equipment supply chain. *Management Science* **49**(12) 1653–1670. URL http://mansci.journal.informs.org/content/49/12/1653.short.

Copeland, Adam, George Hall. 2005. The Response of prices, sales, and output to temporary changes in demand **269**(August 2005) 232–269. doi:10.1002/jae. URL http://www.nber.org/papers/w11870.

Cusumano, Michael A, Akira Aldkira Takeishi. 1991. Supplier relations and management: A survey of Japanese, Japanese-transplant, and U.S. auto plants. *Strategic Management Journal* **12**(8) 563–588. doi:10.1002/smj.4250120802. URL http://doi.wiley.com/10.1002/smj.4250120802.

Eichenbaum, M. S. 1989. Some empirical evidence on the production level and production cost smoothing models of inventory investment. *The American Economic Review* **79**(4) 853–864. URL http://www.nber.org/papers/w2523.

Fair, R. C. 1989. The production-smoothing model is alive and well. *Journal of Monetary Economics* **24**(3) 353–370. URL http://www.sciencedirect.com/science/article/pii/0304393289900263.

Gavirneni, S., R. Kapuscinski, S. Tayur. 1999. Value of information in capacitated supply chains. *Management Science* **45**(1) 16–24.

Gopal, Anandasivam, Manu Goyal, Serguei Netessine, Matthew Reindorp. 2013. The impact of new product introduction on plant productivity in the North American automotive industry. *Management Science* **1909**(25) 1–20.

Goyal, Manu, Serguei Netessine, Taylor Randall. 2006. Deployment of manufacturing flexibility : an empirical analysis of the North American automotive industry. URL http://ssrn.com/abstract=2077659.

Graves, S. C., D. B. Kletter, W. B. Hetzel. 1998. A dynamic model for requirements planning with application to supply chain optimization. *Operations Research* **46**(3-Supplement-3) S35–S49. doi:10.1287/opre.46.3.S35.

Graves, S.C. 1981. A review of production scheduling. *Operations Research* **29**(4) 646–675. URL http://www.jstor.org/stable/170383.

Guajardo, Jose A, Morris A Cohen, Serguei Netessine. 2012. Service competition and product quality in the U . S . automobile industry.

Hall, G. J. 2000. Non-convex costs and capital utilization : A study of production scheduling at automobile assembly plants. *Journal of Monetary Economics* **45**. URL http://www.sciencedirect.com/science/article/pii/S030439320000009X.

Hardle, W., J. Horowitz, J. P. Kreiss. 2003. Bootstrap methods for time series. *International Statistical Review/Revue Internationale de Statistique* 435–459.

Hausman, W. H. 1969. Sequential decision problems: a model to exploit existing forecasters. *Management Science* **16**(2) 93–111.

Heath, D. C., P. L. Jackson. 1994. Modeling the evolution of demand forecasts ITH application to safety stock analysis in production/distribution systems. *IIE transactions* **26**(3) 17–30. URL http://www.informaworld.com/index/779158830.pdf.

Helper, Susan R, Mari Sako. 1995. Supplier Relations in Japan and the United States: Are They Converging? *Sloan Management Review* **36**(3) 77–84. URL http://sloanreview.mit.edu/the-magazine/articles/1995/spring/3636/supplier-relations-in-japan-and-the-united-states-are-they-converging/.

Henningsen, Arne, Jeff D Hamann. 2007. systemfit: A Package for Estimating Systems of Simultaneous Equations in R. *Journal Of Statistical Software* **23**(4) 1–40. URL http://www.jstatsoft.org/v23/i04/.

Holt, C. C., F. Modigliani, J. F. Muth, H. A. Simon. 1960. *Planning production, inventories, and work forces*. Prentice-Hall, Englewood Cliffs, NJ.

Huson, M., D. Nanda. 1995. The impact of just-in-time manufacturing on firm performance in the US. *Journal of Operations Management* **12**(3-4) 297–310.

Kahn, James A. 1987. Inventories and the volatility of production. *The American Economic Review* **77**(4) 667–679.

Keane, Michael P. 2010. Structural vs. atheoretic approaches to econometrics. *Journal of Econometrics* **156**(1) 3–20. doi:10.1016/j.jeconom.2009.09.003. URL http://linkinghub.elsevier.com/retrieve/pii/S0304407609001948.

Kleindorfer, P.R., CH Kriebel, GL Thompson, GB Kleindorfer. 1975. Discrete optimal control of production plans. *Management Science* **22**(3) 261–273. URL http://www.jstor.org/stable/2629640.

Lai, KS. 1991. Aggregation and Testing of the Production Smoothing Hypothesis. *International Economic Review* **32**(2) 391–403. URL http://www.jstor.org/stable/10.2307/2526882.

Lee, H. L., V. Padmanabhan, S. Whang. 1997. Information distortion in a supply chain : the bullwhip effect. *Management science* **43**(4) 546–558. URL http://www.jstor.org/stable/2634565.

Lee, H. L. HL, KC C. So, CS S. Tang. 2000. The value of information sharing in a two-level supply chain. *Management Science* 626–643URL http://mansci.journal.informs.org/content/46/5/626.short.

Lee, Hau L. 2004. The triple-A supply chain. *Harvard Business Review* **82**(10) 102–113. URL http://www.ncbi.nlm.nih.gov/pubmed/15559579.

Lippman, S.A., A.J. Rolfe, H.M. Wagner, J.S.C. Yuan. 1967. Algorithms for optimal production scheduling and employment smoothing. *Operations Research* **15**(6) 1011–1029. URL http://www.jstor.org/stable/168607.

Lucas, R. E. 1976. Econometric policy evaluation: a critique. *Carnegie-Rochester conference series on public policy*, vol. 1. 19–46. URL http://www.econ.umn.edu/~ceyhun/teaching/summer08/48903/lucas1976.pdf.

Miron, J. A., S. P. Zeldes. 1988. Seasonality, cost shocks, and the production smoothing model of inventories. *Econometrica: Journal of the Econometric Society* **56**(4) 877–908. URL http://www.jstor.org/stable/10.2307/1912703.

Moreno, Antonio, Christian Terwiesch. 2011. Pricing and Production Flexibility : An Empirical Analysis of the U . S . Automotive Industry. *Working Paper* 1–31.

Newey, W. K. 1984. A method of moments interpretation of sequential estimators. *Economics Letters* **14**(2-3) 201–206.

Olivares, M., C. Terwiesch, L. Cassorla. 2008. Structural estimation of the newsvendor model: an application to reserving operating room time. *Management Science* **54**(1) 41–55. doi:10.1287/mnsc.1070.0756. URL http://mansci.journal.informs.org/cgi/doi/10.1287/mnsc.1070.0756.

Ramey, V, K West. 1999. Chapter 13 Inventories. *Handbook of Macroeconomics* **1** 863–923. doi:10.1016/S1574-0048(99)10021-1. URL http://linkinghub.elsevier.com/retrieve/pii/S1574004899100211.

Ramey, VA. 1991. Nonconvex costs and the behavior of inventories. *Journal of Political Economy* **99**(2) 306–334. URL http://www.jstor.org/stable/10.2307/2937683.

Ramey, VA, DJ Vine. 2006. Declining volatility in the US automobile industry. *The American economic review* (2000). URL http://www.jstor.org/stable/10.2307/30035001.

Reiss, P C, F A Wolak. 2007. Structural econometric modeling: Rationales and examples from industrial organization. *Handbook of Econometrics* 4277–4415.

Rossi, P.E., G.M. Allenby, R. McCulloch. 2005. *Bayesian statistics and marketing*. John Wiley & Sons, West Sussex, UK.

Simchi-Levi, David, Yao Zhao. 2003. The value of information sharing in a two-stage supply chain with production capacity constraints. *Naval Research Logistics* **50**(8) 888–916. doi:10.1002/nav.10094. URL http://dx.doi.org/10.1002/nav.10094.

Terwiesch, Christian, Z. Justin Ren, Teck H. Ho, Morris a. Cohen. 2005. An empirical analysis of forecast sharing in the semiconductor equipment supply chain. *Management Science* **51**(2) 208–220. doi:10.1287/mnsc.1040.0317.

Veeraraghavan, S, A Scheller-Wolf. 2008. Now or later: a simple policy for effective dual sourcing in capacitated systems. *Operations Research* **56**(4) 850–864. doi:10.1287/opre.1080.0552. URL http://or.journal.informs.org/cgi/doi/10.1287/opre.1080.0552.

West, K.D. 1985. A variance bounds test of the linear quadratic inventory model. URL http://www.nber.org/papers/w1581.