M&SOM MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 9, No. 4, Fall 2007, pp. 457–479
ISSN 1523-4614 | EISSN 1526-5498 | 07 | 0904 | 0457

informs®

DOI 10.1287/msom.1060.0149

In Search of the Bullwhip Effect

Gérard P. Cachon
The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104,
cachon@wharton.upenn.edu

Taylor Randall, Glen M. Schmidt
David Eccles School of Business, University of Utah, Salt Lake City, Utah 84112
{taylor.randall@business.utah.edu, glen.schmidt@business.utah.edu}

The bullwhip effect is the phenomenon of increasing demand variability in the supply chain from downstream echelons (retail) to upstream echelons (manufacturing). The objective of this study is to document the strength of the bullwhip effect in industry-level U.S. data. In particular, we say an industry exhibits the bullwhip effect if the variance of the inflow of material to the industry (what macroeconomists often refer to as the variance of an industry's "production") is greater than the variance of the industry's sales. We find that wholesale industries exhibit a bullwhip effect, but retail industries generally do not exhibit the effect, nor do most manufacturing industries. Furthermore, we observe that manufacturing industries do not have substantially greater demand volatility than retail industries. Based on theoretical explanations for observing or not observing demand amplification, we are able to explain a substantial portion of the heterogeneity in the degree to which industries exhibit the bullwhip effect. In particular, the less seasonal an industry's demand, the more likely the industry amplifies volatility—highly seasonal industries tend to smooth demand volatility whereas nonseasonal industries tend to amplify.

Key words: bullwhip effect; production smoothing; supply chain management; volatility
History: Received: April 29, 2005; accepted: December 6, 2006.

1. Introduction

Lee et al. define the bullwhip effect as "the amplification of demand variability from a downstream site to an upstream site" (2004, p. 1887). In a seminal paper (1997a), these same authors outline four causes of the bullwhip effect and suggest several managerial practices to mitigate its consequences. Their work was motivated, at least in part, by the observation of the bullwhip effect at individual companies: e.g., in the supply chain for diapers, Procter and Gamble noticed that the volatility of demand on its factories was quite high even though it was confident end consumer demand was reasonably stable (Lee et al. 1997a); Hammond (1994) finds amplification in Barilla's pasta supply chain; and Lee et al. (1997b) observe it at a soup manufacturer. Sterman (1989) reports the bullwhip effect when subjects manage a single product in a simulated supply chain (the "beer game"). The bullwhip effect has also been associated with industry-level volatility: Holt et al. (1968) report the bullwhip effect in the television set industry; Anderson et al. (2000) attribute the substantial volatility in the machine tool industry to the bullwhip effect; and Terwiesch et al. (2005) note that the semiconductor equipment industry is more volatile than the personal computer industry. Lee et al. (1997a) relate their work on the bullwhip effect to the extensive literature in macroeconomics that finds industry-level production to be more volatile than demand (e.g., Blanchard 1983, Blinder and Maccini 1991).

Our main objective is to search for the bullwhip effect in recent industry-level data from the U.S. Census Bureau. These monthly data are from January 1992 to February 2006 and cover six retail, 18 wholesale, and 50 manufacturing industries. For each industry in the sample, we measure the volatility of demand imposed on that industry by its downstream customers and the volatility of the inflow of material to that industry, a measure we call "production." We say an industry exhibits the bullwhip effect when the variability of the inflow to the industry (production) is greater than the variability of the outflow from the industry (demand). We review the set of arguments that theorize why an industry (or a firm) may amplify variance and the set of arguments that theorize why an industry (or a firm) may attenuate demand variance. We emphasize that we are testing the net effect of amplification versus attenuation, not whether the theoretical arguments for variance amplification or attenuation have empirical support individually.

Although we define the bullwhip effect in binary terms (an industry exhibits the bullwhip effect or not), we observe considerable heterogeneity in the degree to which industries amplify or dampen variability. Hence, we seek to explain this heterogeneity. The bullwhip effect also implies that upstream echelons of the supply chain should have higher demand volatility than lower echelons. Therefore, we compare demand volatility among retailers, wholesalers, and manufacturers to determine if manufacturers face the most volatility and retailers the least. Finally, awareness of the bullwhip effect has increased over our sample period along with significant changes in information technology and supply chain practices. As a result, we test for trends over time in the volatility of production relative to the volatility of demand.

We begin, in the next section, with a summary of the related literature. The subsequent sections detail our data and explain how we identify the bullwhip effect. Section 5 outlines our hypotheses and §6 describes our analysis. The final section summarizes and discusses our results.

2. Literature Review

The economics literature on supply chain volatility is extensive and generally precedes the work in operations management. However, instead of the bullwhip effect, economists frame their discussion in terms of production smoothing. A firm can smooth its production relative to its sales (i.e., its production is less volatile than sales) by using inventory as a buffer. Such behavior is desirable for a firm if maintaining production at an constant level is less costly than varying the level of production, possibly because the production cost function is convex in the amount produced (i.e., increasing marginal cost) or because it is costly to change the rate of production. For example, suppose a firm faces predictable variability in its demand throughout the year (i.e., seasonality). Production smoothing is then an appropriate strategy: Produce at a reasonably constant rate throughout the year, building inventory during the low season and drawing down inventory during the high season. Production smoothing is also desirable with the combination of seasonality and stochastic shocks (Sobel 1969).[1]

Given that the intuition behind production smoothing is simple and compelling, one would expect to find data indicating production smoothing behavior easily. Yet, much to the surprise of economists, the majority of the empirical evidence finds production more variable than sales. Blanchard (1983) concludes that “in the automobile industry, inventory behavior is destabilizing: the variance of production is larger than the variance of sales.” Blinder (1986) states “the production smoothing model is in trouble. Certain overwhelming facts seem not only to defy explanation within the production smoothing framework, but actually to argue that the basic idea of production smoothing is all wrong.” Miron and Zeldes (1988) conclude “results of our empirical work provide a strong negative report on the production smoothing model.” Eichenbaum (1989) finds “overwhelming evidence against the production-level smoothing model.” Summarizing the literature, Blinder and Maccini (1991) write “the basic facts to be explained are these: ... production is more variable than sales in most industries.” Additional negative findings on the production-smoothing model are reported by Kahn (1987, 1992), Krane and Braun (1991), Mosser (1991), Rossana (1998), and West (1986).[2]

The fact that empirical evidence was not aligning well with the production-smoothing hypothesis

[1] There are other conditions that lead to production smoothing. For example, if production costs are convex in the production rate, then production smoothing is appropriate with stationary and stochastic demand. Abel (1985) shows that it is also desirable even if marginal production costs are constant in the production rate, as long as there is a lead time to produce and excess demand over inventory is lost (i.e., not backordered).

[2] Although most of the published literature focuses exclusively on the U.S. economy, Beason (1993) and Mollick (2004) study Japanese industry-level data and find evidence in support of production smoothing. They conjecture that Japanese firms may be more likely to production smooth due to a better understanding of modern manufacturing techniques and the inability of firms to easily change their labor forces.

motivated economists to explore this conflict between theory and observation. Some argued there were problems with the econometric analysis of production smoothing. For example, Fair (1989) suggests that tests of production smoothing are flawed when they are based on production and when sales are measured in monetary units rather than actual physical units. In support of his conjecture, he finds production smoothing in several industries for which physical data were available. Ghali (1987) argues that tests with seasonally adjusted data are biased against production smoothing and observes production smoothing in seasonally unadjusted data from the cement industry. Unfortunately, to our knowledge, seasonally unadjusted data were not available for a broad set of industries at that time, so it is difficult to test Ghali's seasonality conjecture. Miron and Zeldes (1988) make an attempt to do so by reintroducing seasonality into their data. Nevertheless, they still find strong evidence to reject production smoothing. Reflecting on these econometric critiques, Blinder and Maccini (1991) suggest that while some industries may smooth production, the preponderance of the evidence indicates that most do not.

Others argue that, instead of a problem with the econometric tests of production smoothing, there are problems with the production-smoothing theory, i.e., perhaps firms should rationally amplify. Blinder (1986) offers cost shocks as an explanation for the observed volatility: If production costs vary, then the firm should increase production when costs are cheap and decrease production when costs are expensive. However, mixed results were obtained on the link between production volatility and factor prices (see Blinder 1986, and Miron and Zeldes 1988). Alternatively, production may be more volatile than sales because firms have an incentive to batch their production. This would occur if firms actually operate in a decreasing marginal cost zone of their production function. Ramey (1991) provides some evidence in support of this idea, but others are skeptical of her cost function estimates (e.g., Blinder and Maccini 1991). Blinder (1981) argues that batching occurs because firms face fixed ordering/setup costs and therefore implement (S, s) policies. Caplin (1985) extends the work on (S, s) policies by demonstrating that their properties are preserved under aggregation; that is, the aggregate production of multiple firms implementing (S, s) policies is more volatile than their aggregate sales no matter the correlation structure of demand. Mosser (1991) provides empirical support for the (S, s) policy explanation.

Kahn (1987) does not critique the cost function behind production smoothing but rather the characteristics of demand. He presumes a firm may face first-order autoregressive demand, AR(1) demand for short. If the AR(1) coefficient is positive, then demand is positively correlated over time and production is then more volatile than sales even if production costs are linear in volume. A positive demand shock causes the firm not only to replace the observed demand shock but to increase production further in anticipation of higher future demand.

The operations management literature refines some of these causes of production volatility. Lee et al. (1997a) extend Kahn's work on AR(1) demand to include positive lead times, and Chen et al. (2000) study AR(1) demand with exponential smoothing forecasts. Graves (1999) studies positively correlated demand with a moving average process, and Gilbert (2005) extends the results of Lee et al. (1997a) and Graves (1999) to ARIMA demand. Additional work on the impact of correlated demand on supply chain variability is found in Aviv (2001, 2002, 2003), Gaur et al. (2005b), and Raghunathan (2001). The influence of order batching is studied by Lee et al. (1997a) and Cachon (1999), and Lee et al. (1997a) identify cost shocks in the form of temporary promotions as a contributor to the bullwhip effect. Sterman (1989) adds the misperception of feedback timing on the part of decision makers as an additional cause, and others raise shortage gaming (competitive bidding for scarce capacity) as a potential culprit (see Lee et al. 1997a, Cachon and Lariviere 1999).

However, Baganha and Cohen (1998) recognize that in a multiechelon system with many retailers and one wholesaler, batching at the retail level causes negatively correlated demand for the wholesaler, which causes the wholesaler to smooth production. Consistent with their theory, in seasonally adjusted data they find empirical evidence that retailers amplify demand

and wholesalers do not. (Interestingly, we find the opposite with seasonally unadjusted data: Retailers smooth and wholesalers amplify.)

To summarize, there are reasons a firm or industry may attenuate demand variability (i.e., impose less volatility on its suppliers than is imposed on it by its customers), such as an increasing marginal cost of production combined with predictable seasonality. However, there are also reasons (such as fixed ordering costs and positively correlated demand) to observe amplification. Furthermore, the forces for variance amplification and variance attenuation may coexist. Hence, whether the bullwhip effect is exhibited (i.e., production is more volatile than demand) depends on the relative importance of these factors, which is an issue best resolved via empirical analysis.

3. Data

This section details the data available for our study and our initial adjustments to the data.

Data from January 1992 to February 2006 were obtained from the U.S. Census Bureau and the Bureau of Economic Analysis (BEA). See Boettcher and Gaines (2004) for a description of the Census process. Census reports monthly sales and inventories for retail (U.S. Census Bureau 2006a), wholesale (U.S. Census Bureau 2006b), and manufacturing (U.S. Census Bureau 2006c) industries as well as aggregate series for each of those three levels in the supply chain. (Some series are reported as "shipments" instead of "sales," but for consistency we shall describe all of these series as "sales," i.e., the physical outflow of product.) For some durable goods manufacturing categories, Census also reports new orders received by that industry from its customers (U.S. Census Bureau 2006c). (Census defines durable goods as high-ticket items that last more than 3–5 years.) We refer to orders received as "demand" and note that demand can differ from sales. Census reports both seasonally unadjusted and seasonally adjusted series, but in all cases we use seasonally unadjusted data. For manufacturing inventories, we use the total inventory measure because detailed finished goods inventories are not available for all industries.

Table 1 lists the industries included in our study. To avoid possible double counting, we included only nonoverlapping industries.[3] For all industries in Table 1, both sales and inventory series are available. For some industries, labeled with a # symbol in Table 1, a demand series is also available from Census. For all other industries, we construct a demand series equal to the industry's sales series; that is, we use an industry's sales as a proxy for its demand. This is reasonable when firms carry stock to satisfy customer demand, generally do not stock out, and when customers are not willing to backorder—plausible conditions for retailers but perhaps less so for wholesalers and manufacturers. For those industries where both sales and demand data are available, sales and demand have similar volatility in most cases; in eight industries, the variance of demand is more than twice that of sales, and no industry had the reverse. Because of the nature of their products, we suspect those eight industries operate in a make-to-order fashion and produce customized goods (e.g., aircraft). Orders are received in a volatile fashion, but because customers wait for delivery, firms smooth the flow of their deliveries to customers. As a result, demand is considerably more volatile than sales.

Demand and sales series are adjusted for margin (i.e., multiplied by one minus the margin) to convert them into cost dollar units used in valuing inventories. Monthly margins for retail and wholesale are obtained from Census (U.S. Census Bureau 2006a, b).[4] Monthly margins are not available for manufacturing. Instead, manufacturing demand and sales are adjusted by the margin reported in the 1997 five-year Census (U.S.

[3] In reporting data for manufacturing, Census uses a three-digit coding system, the M3 Series Identification Code. (Some codes are three letters and others are two numbers followed by a letter.) Each three-digit code, such as other durable goods (ODG), subsumes one or more North American Industrial Classification System (NAICS) codes; e.g., ODG is a compilation of NAICS codes 321, 327, 337, and 339. The NAICS system is hierarchical (for example, each three-digit code "337" combines all four-digit codes beginning with 337, such as 3371, 3372, …). The M3 codes are not necessarily hierarchical; e.g., data for one NAICS code may be compiled into data for multiple M3 codes. Thus, we included all manufacturing series such that there are no NAICS codes found in more than one series. There is no overlap in NAICS codes among wholesale and retail industries, so we included all available series.

[4] In some cases, margins are not reported at the same level of NAICS code as sales and inventory data are, so some judgment is made in applying the margin numbers.

Table 1 Amplification Measures for Industry Groups (1992–2005)

	$V[Y]/V[D]$	$V[Y]-V[D]$	*p*-value for equal variance	$V[Y']/V[D']$	$V[Y']-V[D']$	*p*-value for equal variance
Retail industries						
Aggregate retail series	0.50	−0.0056	0.0000	**1.88**	**0.0005**	0.000
Building material and garden equipment and supplies dealers	0.94	−0.0005	0.3511	**1.32**	**0.0005**	0.036
Clothing and clothing accessory stores	0.35	−0.0618	0.0000	**4.63**	**0.0046**	0.000
Food and beverage stores	0.98	−0.0001	0.4494	**1.30**	**0.0001**	0.045
Furniture, home furnishings, electronics, and appliance stores	0.63	−0.0084	0.0015	**5.82**	**0.0034**	0.000
General merchandise stores	0.29	−0.0408	0.0000	**1.41**	**0.0006**	0.013
Motor vehicle and parts dealers	**1.95**	**0.0062**	0.0000	**1.68**	**0.0018**	0.000
Wholesale industries						
Aggregate wholesale series	**1.14**	**0.0006**	0.1943	**1.15**	**0.0002**	0.190
Apparel, piece goods, and notions	**1.24**	**0.0039**	0.0860	**2.68**	**0.0068**	0.000
Beer, wine, and distilled alcoholic beverages	0.57	−0.0101	0.0002	**1.39**	**0.0017**	0.017
Chemicals and allied products	**1.48**	**0.0025**	0.0054	**2.01**	**0.0029**	0.000
Drugs and druggists' sundries	**4.15**	**0.0164**	0.0000	**2.56**	**0.0044**	0.000
Electrical and electronic goods	0.99	0.0000	0.4749	**1.75**	**0.0014**	0.000
Farm product raw materials	**3.45**	**0.0285**	0.0000	**2.26**	**0.0076**	0.000
Furniture and home furnishings	**1.45**	**0.0027**	0.0083	**2.28**	**0.0029**	0.000
Grocery and related products	**1.39**	**0.0013**	0.0162	**1.65**	**0.0007**	0.001
Hardware, and plumbing and heating equipment and supplies	**1.17**	**0.0009**	0.1593	**1.76**	**0.0018**	0.000
Lumber and other construction materials	**1.11**	**0.0009**	0.2429	**1.30**	**0.0009**	0.047
Machinery, equipment, and supplies	**1.24**	**0.0019**	0.0812	**1.73**	**0.0019**	0.000
Metals and minerals, for example, petroleum	**1.50**	**0.0031**	0.0047	**1.75**	**0.0021**	0.000
Miscellaneous durable goods	**1.15**	**0.0010**	0.1905	**1.96**	**0.0024**	0.000
Miscellaneous nondurable goods	**1.42**	**0.0025**	0.0120	**2.32**	**0.0029**	0.000
Motor vehicle and motor vehicle parts and supplies	**1.11**	**0.0008**	0.2513	**2.44**	**0.0033**	0.000
Paper and paper products	**1.67**	**0.0034**	0.0005	**1.80**	**0.0018**	0.000
Petroleum and petroleum products	**1.35**	**0.0013**	0.0249	**1.63**	**0.0012**	0.001
Professional and commercial equipment and supplies	**1.07**	**0.0007**	0.3363	**1.34**	**0.0009**	0.030
Manufacturing industries						
Aggregate manufacturing series	0.55	−0.0028	0.0001	0.79	−0.0001	0.068
Apparel	0.57	−0.0057	0.0001	**1.40**	**0.0006**	0.014
Audio and video equipment manufacturing	0.86	−0.0034	0.1693	**1.89**	**0.0069**	0.000
Automobile manufacturing	0.90	−0.0071	0.2371	**1.07**	**0.0004**	0.341
Battery manufacturing	**1.06**	**0.0020**	0.3466	**2.26**	**0.0096**	0.000
Beverage manufacturing	**3.04**	**0.0080**	0.0000	**9.18**	**0.0065**	0.000
Communications equipment manufacturing, defense#	0.93	−0.0189	0.3207	**1.13**	**0.0235**	0.222
Communications equipment manufacturing, nondefense#	0.35	−0.0530	0.0000	**1.04**	**0.0008**	0.406
Computer storage device manufacturing	0.20	−0.3519	0.0000	**2.17**	**0.0250**	0.000
Construction machinery manufacturing#	0.73	−0.0061	0.0213	0.50	−0.0076	0.000
Dairy product manufacturing	0.85	−0.0005	0.1439	**1.57**	**0.0006**	0.002
Electric lighting equipment manufacturing	0.43	−0.0155	0.0000	0.66	−0.0031	0.003
Electrical equipment manufacturing#	0.70	−0.0048	0.0107	**1.14**	**0.0009**	0.200
Electromedical, measuring, and control instrument manufacturing#	0.48	−0.0201	0.0000	**1.05**	**0.0005**	0.373
Electronic computer manufacturing#	0.39	−0.0387	0.0000	0.71	−0.0050	0.013
Fabricated metal products#	**1.01**	**0.0000**	0.4755	0.91	−0.0001	0.279
Farm machinery and equipment manufacturing	0.88	−0.0028	0.2038	**1.32**	**0.0024**	0.035
Ferrous metal foundries#	**1.21**	**0.0040**	0.1075	0.73	−0.0016	0.021
Furniture and related products#	**1.10**	**0.0004**	0.2677	0.97	−0.0001	0.417
Grain and oilseed milling	**2.90**	**0.0066**	0.0000	**2.01**	**0.0020**	0.000
Heavy duty truck manufacturing	**1.13**	**0.0017**	0.2082	**1.10**	**0.0005**	0.262
Household appliance manufacturing#	0.69	−0.0039	0.0093	0.63	−0.0026	0.001
Industrial machinery manufacturing#	0.23	−0.0404	0.0000	0.36	−0.0182	0.000

Table 1 (cont'd.)

	$V[Y]/V[D]$	$V[Y]-V[D]$	p-value for equal variance	$V[Y']/V[D']$	$V[Y']-V[D']$	p-value for equal variance
Iron and steel mills and ferroalloy and steel products manufacturing#	0.81	−0.0017	0.0836	**1.07**	**0.0004**	0.331
Leather and allied products	0.79	−0.0033	0.0589	**2.04**	**0.0052**	0.000
Light truck and utility vehicle manufacturing	0.97	−0.0023	0.4228	**1.02**	**0.0002**	0.447
Material handling equipment manufacturing#	0.33	−0.0228	0.0000	0.39	−0.0107	0.000
Meat, poultry, and seafood product processing	**1.08**	**0.0002**	0.3187	**1.19**	**0.0002**	0.128
Metalworking machinery manufacturing#	0.79	−0.0049	0.0658	**1.73**	**0.0060**	0.000
Mining, oil, and gas field machinery manufacturing#	**2.10**	**0.0262**	0.0000	**1.86**	**0.0162**	0.000
Miscellaneous products#	0.65	−0.0038	0.0024	**2.29**	**0.0016**	0.000
Nonmetallic mineral products	0.79	−0.0011	0.0658	**1.24**	**0.0003**	0.081
Other computer peripheral equipment manufacturing	0.31	−0.1139	0.0000	**1.02**	**0.0003**	0.447
Other electronic component manufacturing#	0.69	−0.0048	0.0084	0.90	−0.0010	0.249
Paint, coating, and adhesive manufacturing	**1.47**	**0.0022**	0.0067	**1.65**	**0.0011**	0.001
Paperboard container manufacturing	**1.40**	**0.0009**	0.0146	**2.58**	**0.0016**	0.000
Pesticide, fertilizer, and other agricultural chemical manufacturing	**0.66**	**−0.0128**	0.0036	**1.28**	**0.0021**	0.054
Petroleum and coal products	2.95	0.0062	0.0000	**3.86**	**0.0054**	0.000
Pharmaceutical and medicine manufacturing	**2.86**	**0.0143**	0.0000	**4.96**	**0.0147**	0.000
Photographic equipment manufacturing#	**1.54**	**0.0111**	0.0028	**1.79**	**0.0084**	0.000
Plastics and rubber products	0.99	0.0000	0.4797	**1.83**	**0.0008**	0.000
Printing	**1.59**	**0.0017**	0.0015	**3.26**	**0.0016**	0.000
Pulp, paper, and paperboard mills	**1.20**	**0.0003**	0.1183	**1.45**	**0.0005**	0.008
Search and navigation equipment manufacturing, defense#	0.15	−0.2958	0.0000	0.21	−0.1602	0.000
Search and navigation equipment manufacturing, nondefense#	**1.23**	**0.0245**	0.0889	**1.48**	**0.0353**	0.006
Textile products	**1.12**	**0.0007**	0.2393	**2.40**	**0.0014**	0.000
Textiles	0.57	−0.0049	0.0001	**1.14**	**0.0002**	0.197
Tobacco manufacturing	**3.09**	**0.0839**	0.0000	**3.08**	**0.0657**	0.000
Transportation equipment#	0.55	−0.0214	0.0001	0.25	−0.0092	0.000
Ventilation, heating, air conditioning, and refrigeration#	0.79	−0.0026	0.0635	0.75	−0.0014	0.033
Wood products	**1.26**	**0.0010**	0.0673	**2.00**	**0.0009**	0.000

Notes. $V[Y]$ = Variance of production; $V[D]$ = Variance of demand; $V[Y']$ = Variance of seasonally adjusted production; $V[D']$ = Variance of seasonally adjusted demand. Bold indicates when the bullwhip is present (when the ratio is greater than 1).
Industry for which demand and sales data are available. In all other industries, sales is used as a proxy for demand.

Census Bureau 2005).[5] Demand, sales, and inventory series are also price-index adjusted so that changes over time reflect real valuations. U.S. BEA (2006) gives implicit price deflators for retail, wholesale, and manufacturing. In applying the price deflators to manufacturing, there are some direct matches, but some judgment is needed as to which BEA deflator should be used for which Census code. However, for retail and wholesale there is a one-to-one correspondence between the categories of BEA price deflators and the Census data. Discussion from here onward is based on price and margin-adjusted data. Our data are available on the *Manufacturing & Service Operations Management* website.

For each industry i, we use its sales and inventory series to evaluate an imputed production series: Production in month t, Y_{it}, is evaluated as

$$Y_{it} = S_{it} + (I_{it} - I_{it-1}), \qquad (1)$$

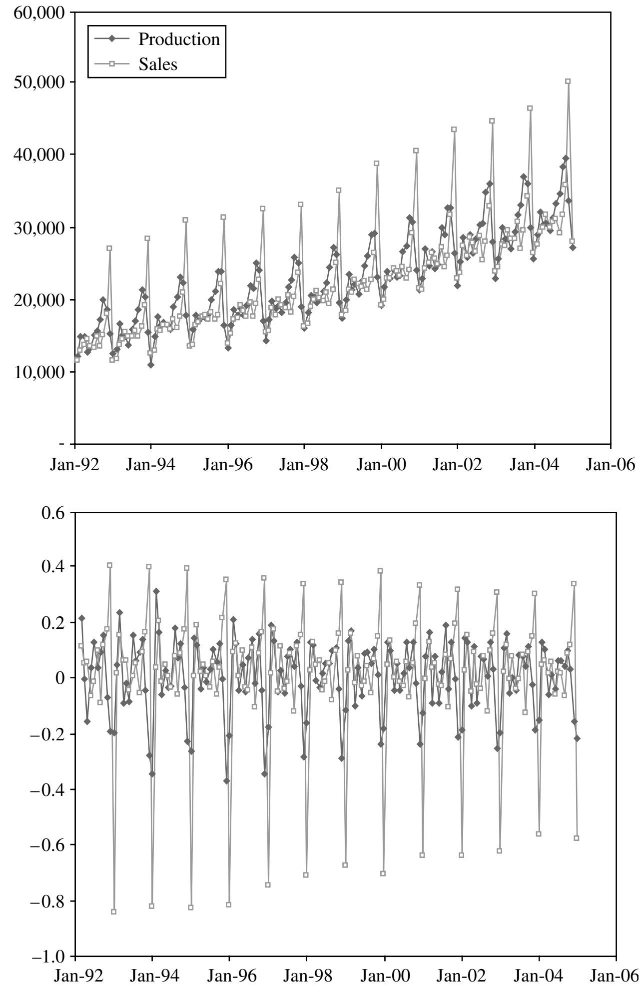
[5] Margins reported by the five-year Census are given by NAICS codes rather than M3 categories, but we agglomerated the data to fit the M3 categories. Though Census separates several M3 categories into defense and nondefense categories in the monthly reporting, it does not do so in the five-year data so we assume defense and nondefense have the same margin for a given NAICS code. 2002 reported margins are nearly equivalent to the 1997 margins, and the 1992 margins could not be determined precisely, as Census changed from the SIC coding system to the NAICS coding system in 1997. Margins are calculated as (value added − production wages)/(value of shipments), which is slightly preferred over [(value of shipments) − (materials purchases + production wages)]/(value of shipments), but the two methods differed by only 0.2% on average.

where S_{it} and I_{it} are the sales and inventory for industry i in month t. (Recall that S_{it} is actually the cost of sales, i.e., margin-adjusted sales.) For retail and wholesale, we interpret production to be the inflow of material to the industry, whereas production for manufacturers represents both the inflow of raw materials and components from suppliers and the industry's own production to convert those inputs into finished goods. In other words, the term "production" should be interpreted broadly—production refers to the net flow of material through an industry, not just the industry's internal processes that convert raw materials to finished goods.

The relationship (1) makes clear why it is necessary to margin adjust the sales series: The production series is evaluated from both the sales and inventory series, so they should be measured in comparable units. Furthermore, it is straightforward to show that evaluating production with nonmargin-adjusted sales will result in overestimation of the production variances. Price adjustment of both series is done to ensure that both series are measured in comparable units over time.

We apply two additional adjustments to each series. We log and first difference each series: Each series $X = \{X_1, X_2, \ldots, X_n\}$ is converted into the following series: $\{\ln(X_2) - \ln(X_1), \ldots, \ln(X_n) - \ln(X_{n-1})\}$. To explain the motivation for these adjustments, consider Figure 1, production and demand for general merchandisers (a retail industry), and Figure 2, production and demand for nondefense communications equipment (a manufacturing industry), or "telecom" for short. It is apparent from the figures that both industries are trending, and the Dickey-Fuller test suggests that most of our series possess a unit root (i.e., they are random walks).[6] As a result, the variances of these series depend on the length of the time horizon, which is undesirable.[7] Furthermore, because firms are unable to sustain permanent deviations in their long-run average production relative to their sales (i.e., they share the same stochastic trend), the variance of the two series converges as the time horizon increases.[8] Hence, we wish to remove the

Figure 1 General Merchandise Stores Production and Sales (Margin and Price Adjusted) (Top Graph) and First Differences of Logged Production and Sales (Bottom Graph)

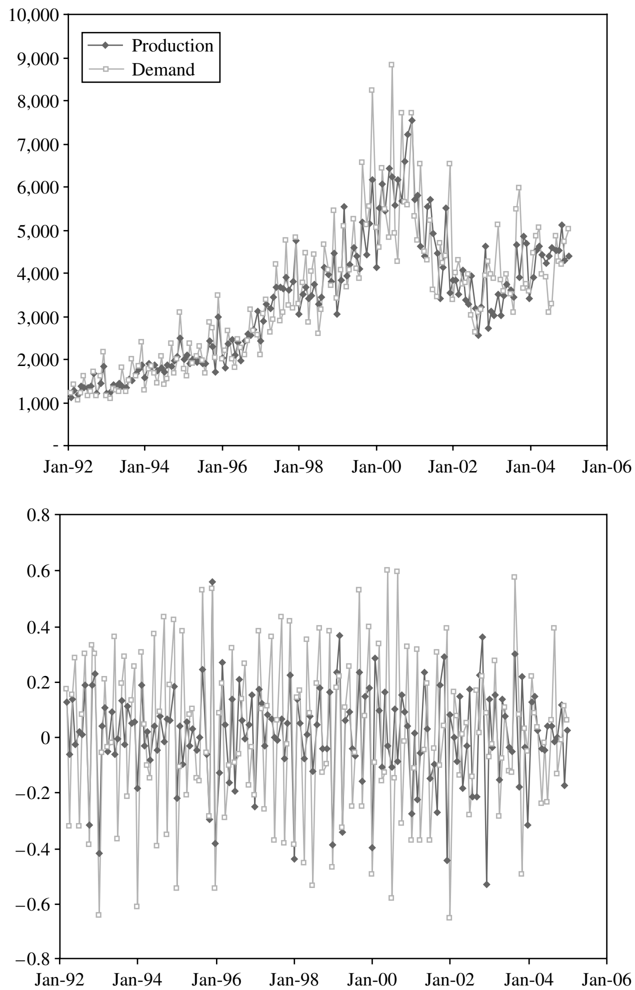


[6] To test for the presence of a unit root in each series, we take as our null hypothesis that the series contains a unit root and then apply the Dickey-Fuller test to determine whether we can reject our null (see Hamilton 1994 for details on this test). Results (available from the authors) indicate that most series fail to reject the null.

[7] Consider a simple unit root process $x_i = x_{i-1} + \varepsilon_i$, where ε are iid shocks with variance σ^2. If we assume $x_0 = 0$, then $V[x_i] = i\sigma^2$, where $V[\,]$ is the variance operator, because each shock has a permanent effect. Given that $V[x_i]$ is increasing in i, the estimated variance of the series $\{x_1, \ldots, x_n\}$ will depend on the length of the series. First differencing $\Delta x_i = x_i - x_{i-1} = \varepsilon_i$ results in a series with constant variance.

[8] Consider two unit root series that share the same stochastic trend: $x_i = \mu_i + \varepsilon_i$, $y_i = \mu_i + \delta_i$, $\mu_i = \mu_{i-1} + \omega_i$, where ε, δ, ω are iid shocks with finite variances σ_ε^2, σ_δ^2, and σ_ω^2, respectively. Assuming $\mu_0 = 0$, then the variances of x_i and y_i are $V[x_i] = i\sigma_\omega^2 + \sigma_\varepsilon^2$ and $V[y_i] = i\sigma_\omega^2 + \sigma_\delta^2$. Thus, $\lim_{i\to\infty} V[x_i] = \lim_{i\to\infty} V[y_i]$ so the variances of the two series become indistinguishable for long series.

Figure 2 **Nondefence Communications Equipment Manufacturing Production and Demand (Margin and Price Adjusted) (Top Graph) and First Differences of Logged Production and Demand (Bottom Graph)**



stochastic trend from each series and focus on the variation of each series about that trend.

The standard approach to detrend a random walk is to apply the first difference operator to each series Δ, $\Delta X_t = X_t - X_{t-1}$, where X is either demand or production. Because we log each series before first differencing, the adjusted series is approximately the percentage change between observations in each series: $\ln(X_i) - \ln(X_{i-1}) \approx (X_i - X_{i-1})/X_{i-1}$. We repeated our analysis with unlogged data (i.e., just first-differenced series) and obtained qualitatively similar results, with one exception (see Footnote 19). The lower graphs in Figures 1 and 2 display the adjusted series. Neither series appears to exhibit a trend. Hence, we are now able to compare variances.

Each production series after first differencing ranges from March 1992 to February 2006. In our analysis, we refer to year t as the 12 months from March of year t to February of year $t+1$. Hence, we have 14 years of data for each series.

To summarize, for each industry we have series measuring two physical flows: the inflow of material, which we call production, and the outflow of material, which we call sales. For some durable goods manufacturers, we also have a demand series, which is an information flow into the industry. For all other industries, we take its sales series to also be its demand series. Demand and sales series are margin adjusted. Demand, sales, and inventory series are price adjusted. To eliminate long-run trends, demand and production series are logged and first differenced.

4. Identifying the Bullwhip Effect

We take two approaches to search for the bullwhip effect. The first measures the amount of volatility an industry contributes to the supply chain: An industry faces volatile demand from its customers, then imposes its own volatility on its suppliers. We say the bullwhip effect is exhibited by an industry when the variance of its production is greater than the variance of its demand, i.e., either if its amplification ratio is greater than one,

$$\textit{Amplification ratio} = \frac{V[\textit{Production}]}{V[\textit{Demand}]}, \tag{2}$$

or if its amplification difference is positive

$$\textit{Amplification difference} = V[\textit{Production}] - V[\textit{Demand}], \tag{3}$$

where $V[\cdot]$ is the variance operator. Recall that for many industries we use sales as a proxy for demand. We suspect this biases our analysis in favor of higher amplification measures: If a firm is able to backlog its demand, sales will be less volatile than demand. As a result, our amplification estimates are biased higher than their true values.[9] We include both measures in

[9] For those industries where there are data on both orders received and shipments, in some cases (such as boats, aircraft, and various

our analysis because there is little theory to suggest one should be preferred over the other.[10]

We emphasize that the amplification measures (2) and (3) only inform us whether the forces to amplify demand are stronger or weaker than the forces to attenuate demand. With our data, we are unable to test whether the theoretical causes of the bullwhip effect are indeed causing variance amplification. For example, we suspect that fixed ordering costs are present in all our industries to some extent (though we have no direct evidence of this), but we cannot determine whether those fixed ordering costs are leading firms to amplify demand. Instead, we are testing whether the combined forces to amplify are sufficiently stronger than the forces to attenuate demand, in which case the industry exhibits the bullwhip effect in the form of its production variance being greater than its demand variance.

Instead of the variance of production, the amplification measures (2) and (3) could be defined in terms of the variance of an industry's orders to its suppliers. If that were done, the amplification measures would include two information flows: the demands imposed on an industry and the demands the industry imposes upstream (its orders to its suppliers). Unfortunately, industry order data are generally not available. Nevertheless, we feel that our measures are reasonable. The variance of an industry's production is likely to be a good proxy for the variance of an industry's orders, in which case our measures would be comparable to measures based on industry orders; that is, the information outflow from an industry is likely to be highly correlated with the physical flow into the industry. (Recall that the production measure reflects inflow of material to an industry, which includes inbound deliveries as well internal production processes.)

For example, Hammond (1994) reports order, sales, and inventory series for a single pasta product at a single location. We evaluated the imputed production as in Equation (1) from those sales and inventory series. The correlation between the reported order series and the imputed production series is 0.99, and the variances of their logged and first-differenced series are 1.28 and 1.25, respectively, suggesting that, at least in this one instance, production and orders are quite similar.[11] Apparently Lee et al. (1997a) share a similar view, because they cite the empirical economics literature as evidence of the bullwhip effect: That literature compares the volatility of an industry's production with the volatility of its sales (which is often our proxy for demand). One could even argue that amplification ratios are more informative if they include the variance of production rather than the variance of orders. Order volatility is generally not costly per se, but rather it is costly if it induces volatility in physical flows. For example, if a retailer submits volatile orders to a wholesaler but the wholesaler nevertheless ships in a smooth fashion (either because it chooses to do so or because the retailer allows it to do so), then the operational consequence of that order volatility is not severe.[12]

Note that the production and demand series used to construct the amplification measures are not seasonally adjusted. We have two reasons for working with seasonally unadjusted data. First, firms must produce to meet demand, not seasonally adjusted demand. Predictable variation is operationally inconsequential only if a firm has a constant marginal production cost and incurs no cost to change its production rate. Such a production function is

defense products) the variability in orders received is dramatically higher than for actual shipments. In no case do we find the variability in shipments to be dramatically higher than in orders (in a few cases it is slightly higher).

[10] As is shown in the next section, in the AR(1) model studied by Lee et al. (1997a), if the lead time is zero, then the amplification difference is linear in the estimated AR(1) terms, whereas the amplification ratio is not, which suggests favoring the difference measure over the ratio measure in the econometric specification. However, there are many other models that exhibit the bullwhip effect and production smoothing, and it is not clear which of the two measures is preferred. Hence, we feel it is best to include both measures. Fortunately, our qualitative results between them are similar, which suggests that our results are robust to the actual measure implemented.

[11] We were unable to obtain the original raw data so we estimated the data from the published graphs, which clearly introduces some measurement error. Thus, we view these two series as being nearly identical.

[12] It is possible that order volatility could be costly (in the sense of translating into physical volatility) if orders are used for forecasting purposes: Higher-order volatility leads to less-accurate forecasts, and less-accurate forecasts lead to more volatile physical flows.

unlikely. Second, seasonality provides a strong motivation to smooth production. For example, as discussed in the introduction, general merchandisers (Figure 1) can predict the annual end-of-year spike in demand. Given capacity constraints on their logistics (receiving inbound shipments, warehousing, store deliveries, shelf restocking, etc.), it is prudent for them to smooth production relative to the predictable demand spike. Thus, seasonally adjusting the series removes a major reason for demand attenuation, thereby leaving only reasons for demand amplification.[13] Given that our focus is to measure the relative strengths of these two forces, a fair comparison should be based on seasonally unadjusted data. Consistent with this belief, we anticipate that an industry's amplification measures should increase if they are evaluated with seasonally adjusted production and demand. Recall that the economics literature generally works with seasonally adjusted data primarily because seasonally unadjusted data were not available at the time those studies were conducted.

Our second approach to identify the bullwhip effect is to compare demand volatility at different levels of the supply chain. If each level of a supply chain exhibits the bullwhip effect, then we should observe that the variance of retail demand is less than the variance of wholesale demand, which in turn is less than the variance of manufacturing demand. Note that we do not construct explicit linear supply chains. For example, although it might be tempting to compare the demand volatility of "textiles" (manufacturing), "apparel" (manufacturing), "apparel, piece goods, and notions" (wholesale), "clothing and clothing accessories stores" (retail), and "general merchandisers" (retail), the outflow of material from each of those industries is not limited to just one other industry, nor are the inflows likely to come from only one industry: Apparel is sold in apparel retail stores as well as general merchandisers, but general merchandisers sell more than apparel and textiles are used in apparel but have other applications as well.[14] Hence, because supply chains are more like supply webs, direct comparisons are problematic. As a result, we merely make comparison of demand volatility at these three different levels.

We also recognize that the flows between our three levels of the supply chain are not necessarily equal. For example, some manufacturers may sell to wholesalers that then sell to different manufacturers that sell a final consumer good through wholesalers and retailers. Some manufacturers may sell their goods directly to consumers, thereby bypassing retailers, and some retailers may be vertically integrated, so their products do not show up in any manufacturing industry. Nevertheless, to the extent that retailers tend to sell directly to consumers and manufacturers tend to sell through intermediaries (wholesalers and retailers), our analysis provides some information on how volatility differs across the supply chain.

5. Variation in the Amplification Measures

We seek to explain heterogeneity in our amplification measures (2) and (3) across industries and across time.

Hypothesis 1. *The amplification ratio and the amplification difference are decreasing in the predictable seasonality ratio.*

Based on arguments given in the previous section, we expect an industry's amplification measures to decrease as the proportion of variability attributable to seasonality increases.[15] We measure the degree of

[13] There is a motivation to production smooth even with stationary and stochastic demand, but that motivation is clearly not as strong as the motivation to production smooth, with predictable variation in demand.

[14] If one were to make these comparisons, the pattern that would emerge is opposite to the bullwhip effect; that is, demand variance decreases as the level of the supply chain increases. The variances of first differenced and logged demand are as follows: clothing stores = 0.0960; general merchandisers = 0.0580; apparel wholesalers = 0.0166; apparel manufacturers = 0.0133; and textile manufacturers = 0.0058.

[15] We take seasonality as the primary source of predictable demand. Another approach is to build a forecasting model that includes seasonality and other variables such as lagged sales and macroeconomic variables, among others. In addition, it is possible that the industry has access to valuable forecasting information that is unavailable to us. Hence, our seasonality ratio is likely to underestimate the amount of predictable demand.

seasonality as follows:

$$\textit{Seasonality ratio} = \frac{V[\textit{Demand}] - V[\textit{Seasonally adjusted demand}]}{V[\textit{Demand}]}. \quad (4)$$

The variance of seasonally adjusted demand is the variance of the residuals from a regression on demand with 11 monthly dummy variables.[16] As can be seen in Figures 1 and 2, with general merchandisers there is a significant amount of seasonality, whereas with telecom there is less seasonality. (The seasonality ratio for general merchandisers is 0.98 while it is 0.60 for telecom.)

Hypothesis 2. *The amplification ratio and the amplification difference are increasing in the variance of the first-differenced price index.*

Lee et al. (1997a) illustrate how trade promotion pricing can lead to the bullwhip effect, and Blinder (1986) demonstrates that cost shocks increase the volatility of production. Previous studies have attempted to measure cost shocks through volatility in factor prices (e.g., labor costs, interest rates, commodity prices, etc.). It is often difficult to know the precise set of factor inputs to an industry, so we take a different approach. We use the variance of an industry's first-differenced price index as a proxy for both promotion pricing and cost shocks. In the presence of sticky pricing, it is possible that output prices (which we are using) are only a weak proxy for input prices, but we suspect that they are nevertheless positively correlated.

Hypothesis 3. *The amplification ratio and the amplification difference are increasing in the persistence of demand shocks.*

Lee et al. (1997a) study the bullwhip effect in a single-item, periodic-review inventory model with constant lead time l and demand D following an AR(1) process:

$$D_t = \phi D_{t-1} + \varepsilon_t, \quad (5)$$

where ε are iid shocks with variance σ^2. (Kahn 1987 studies a similar model with zero lead time.) From Equation (3.5) in Lee et al. (1997a), our amplification measures in their model are

$$\frac{V[\textit{Production}]}{V[\textit{Demand}]} = 1 + \frac{2\phi(1-\phi^{l+1})(1-\phi^{l+2})}{1-\phi}$$

$$V[\textit{Production}] - V[\textit{Demand}] = \frac{2\phi(1-\phi^{l+1})(1-\phi^{l+2})}{(1+\phi)(1-\phi)^2}\sigma^2.$$

Equation (5) implies no seasonality in demand, so we estimate ϕ and σ^2 with seasonally adjusted demand series. (See Hamilton 1994 for details on estimating an AR(1) model.) Across all industries, we obtain $\phi \in [-0.67, -0.04]$. Lee et al. (1997a) demonstrate that the bullwhip effect is present when $\phi > 0$ and not present otherwise. (Lee et al. (1997a) compare the variance of orders with the variance of demand, but because lead times are constant, the variance of orders is equivalent to the variance of production, i.e., the inflow of material.) Although the amplification ratio is not monotone in ϕ and neither measure is linear in ϕ, both amplification ratios are approximately increasing and linear in ϕ for the range of our estimates. Hence, we conjecture that the amplification ratio should be increasing in ϕ and the amplification difference should be increasing in $\phi\sigma^2$. (In fact, for $l = 0$, the amplification difference is exactly linear and increasing in $\phi\sigma^2$.)

Hypothesis 4. *The amplification ratio and the amplification difference are decreasing over time.*

Our data span a 14-year period in the U.S. economy that includes both significant growth (most of the 1990s) as well as a period of contraction around 2001. During this period, there were also significant improvements in information technology and supply chain management as well as a significant increase in international trade. Furthermore, recent work in the macroeconomic literature suggests that there has been a decrease in volatility in the U.S. economy over the last 50 years (Blanchard and Simon 2001, McConnell and Perez-Quiros 2000). As a result, we test if there are identifiable time trends in the data.

The industries in our sample vary considerably in size, where we take the log of average (unlogged) sales as the proxy for size. Large industries probably include more firms and sell a broader array of products than small industries. Hence, we expect an

[16] An alternative to using regression to remove seasonality is to take 12th differences of the time series, i.e., $\Delta_{12}X_t = X_t - X_{t-12}$. Results using that approach are qualitatively similar to the results shown in Table 1.

industry's variance of demand to decrease in its size, and a similar argument applies to the variance of an industry's production. In fact, the Spearman correlation coefficients between our size covariate and the variance of production and the variance of demand are −0.41 and −0.44, respectively, both of which are significant at $p < 0.01$. However, it is not clear how industry size influences the amplification measures (e.g., size decreases the numerator and denominator of the ratio, so its net effect is ambiguous). Thus, in our regressions we include industry size as a control variable without stating a specific hypothesis on its influence on the amplification measures.

6. Analysis

We divide our analysis into two sections. The first section explores the prevalence of the bullwhip effect and the second seeks to explain variation in the intensity of the bullwhip effect across industries and across time. Section 6.3 summarizes the results from our analysis.

6.1. Prevalence of the Bullwhip Effect

Table 1 reports our amplification measures (2) and (3) for retail, wholesale, and manufacturing industries from 1992 to 2005 (recall that each of those 14 years begins in March of that year and ends in February of the following year). A ratio larger than one or a positive difference indicates that an industry's production variance is greater than its demand variance, in which case we say the bullwhip effect is exhibited by that industry. A ratio less than one or a negative difference indicates production smoothing. (To emphasize a point, these amplification measures are only able to test for the relative strength of the forces to amplify and the forces to attenuate demand. Even if production smoothing dominates, the causes of the bullwhip effect may be present in the industry, just not sufficiently strong to overcome the causes of production smoothing. Thus, we do not test whether the theoretical causes of the bullwhip effect have empirical support, e.g., we are not testing whether fixed ordering costs lead to amplification.)

For ease of exposition, we only refer to the results using amplification ratios as the results using differences yield identical conclusions. To provide a bridge between our work and previous studies, we also report (in Columns 4 and 5 of the table) the amplification measures evaluated with seasonally adjusted production and demand: As with the seasonality ratio (4), we seasonally adjust those series by regressing the first-differenced data on indicator variables for each month, and then the variances of the residuals are estimated.[17] Columns 3 and 6 of Table 1 provide p-values for the Brown-Forsythe test of homogeneous variance (Brown and Forsythe 1974).[18] We consider differences statistically significant at the $p < 0.10$ level in the following discussion.

In the first column of Table 1, most retail industries (83%) have ratios less than one, indicating a propensity to production smooth rather than to amplify demand variability.[19] In contrast, 89% of the wholesale industries have ratios greater than one, indicating a propensity to amplify. The results for manufacturing are mixed: 60% production smooth and 40% amplify. Across all industries, 50% (37 industries) amplify and 50% (37 industries) production smooth. Significant differences between the production and demand variances occur in 65% of the industries that amplify (24 industries) and 73% of industries that production smooth (27 industries).[20]

[17] We repeated this analysis with the production and demand series adjusted with a common seasonality. It can be shown that the amplification difference is the same with individually or commonly adjusted series. The amplification ratios and the p-values are slightly different, but the results are qualitatively identical. We conclude that our results in Table 1 are robust to whether the series are individually deseasonalized (as reported in the table) or commonly deseasonalized.

[18] For each series, we evaluated variances by each of the 14 years. The test compares the set of 14 production variances to the set of 14 demand variances.

[19] "Motor vehicles and parts dealers" is the only retail category that amplifies. Unlike most retailers, auto dealers have little control on the inflow of product because auto manufacturers are able to push their inventory onto dealers. This may explain why the inflow of product to auto dealers is more variable than auto demand is.

[20] Interestingly, food and beverage retailing, which is presumably the segment selling diapers, pasta, and beer, is nearly variance neutral (amplification ratio 0.98). Beer, wine, and alcoholic beverage wholesalers actually have the lowest amplification ratio among wholesalers (0.57); beverage manufacturing has the second highest amplification ratio among manufacturers (3.04). We note that industrial machinery manufacturing and metalworking machinery manufacturing, which probably include machine tools, are production smoothers (0.23 and 0.79).

Different results are observed, as we expect, when the amplification ratio is evaluated with deseasonalized production and demand: All retail industries, all wholesale industries, and 74% of manufacturing industries amplify in these data (Table 1, Column 4). Across all industries, 82% (61 industries) amplify and 81% (60 industries) show significant differences between production variance and demand variance.

To compare the variability of demand across the three levels of the supply chain, we first examine differences in the mean variance of demand for the aggregate data series for retail, wholesale, and manufacturing. These are calculated based on yearly variances of demand for each year from 1992 to 2005. The mean variances of demand are 0.012, 0.005, and 0.007 for retail, wholesale, and manufacturing aggregate series, respectively. Each mean is statistically different from each other mean at the $p < 0.01$ level, but the trend in variance is inconsistent with the bullwhip hypothesis: Retail has the highest variance, followed by manufacturing and then wholesaling.[21] Interpreting these results in conjunction with the previous discussion, the implication is that retailers smooth customer demand and then wholesalers inject more variability back into the demand signal so that manufacturers face an intermediate amount of demand variability.

Our second method of comparing demand variability in different levels of the supply chain examines mean differences between the demand variation across all retail, wholesale, and manufacturing industries, where we weigh each industry equally. The mean variances of demand are 0.033, 0.008, and 0.046 for retail, wholesale, and manufacturing industries, respectively: There is a significant difference between wholesale and manufacturing but no significant differences among the other levels of the supply chain. Again, these tests are inconsistent with the bullwhip effect, because wholesale industries have the lowest variance of demand.[22] A concern is that we have a limited number of retail industries and these retail industries are generally more aggregated than the manufacturing industries (i.e., they have higher sales volumes). To address this concern, we obtained price- and margin-adjusted demand data for 24 nonoverlapping retail industries. (Inventory data are not available for these series, so we are unable to use them to evaluate amplification ratios.) The mean variance in demand from this retail sample is 0.038. However, we note no significant difference between the disaggregated retail series and wholesale and manufacturing series. Thus, our result appears to be robust to aggregation across retailers.

6.2. Variation in the Bullwhip Effect

Across our sample of industries, there is considerable variation in the amplification measures. This section tests our hypotheses regarding the sources of that variation across industries and across time. Table 2 provides descriptive statistics for our study variables. We divide the table into four panels to show descriptive statistics for the entire data set, retail, wholesale, and manufacturing industries. The mean seasonality ratio for the sample is 0.64, with retailer industries having the highest degree of seasonality (0.87) and wholesale industries the lowest (0.61). Price variance is highest in wholesale industries (4.72) and lowest in retail industries (0.16). The estimated AR(1) coefficients are all negative, which, according to the analytical models in Kahn (1987) and Lee et al. (1997a), should result in amplification ratios less than one. Nevertheless, as mentioned earlier, the amplification ratios may still be increasing in the AR(1) coefficient.

Table 3 provides Spearman correlation coefficients for the panel data.[23] The correlation tables provide

[21] For completeness, we report the same test with seasonally adjusted data. With those data, the mean variances of demand are 0.0006, 0.0020, and 0.0007 for retail, wholesale, and manufacturing aggregate series, respectively. There are statistical differences between each series. These data are also inconsistent with the bullwhip effect, given that wholesale has the highest demand variance and manufacturing demand variance is not significantly higher than retail demand variance.

[22] For seasonally adjusted data, the mean variances of demand are 0.002, 0.003, and 0.016 for retail, wholesale, and manufacturing industries respectively. We note a significant difference between manufacturers and the other categories. This pattern is consistent with the bullwhip effect. However, it is not clear how to interpret the relevance of seasonally adjusted demand. In particular, seasonally adjusted variances are much smaller than unadjusted variances, so measurement error can have a greater impact on the amplification ratios.

[23] Spearman correlation is a nonparametric correlation technique. We use the Spearman correlation because of the small sample sizes of the wholesale and retail subsamples.

Table 2 **Descriptive Statistics of Study Variables (1992–2005)**

	Mean	Std. dev.	10th	50th	90th
Panel A: Entire data set ($N = 74$)					
Amplification ratio	1.160	0.789	0.352	1.000	2.479
Amplification difference	−0.012	0.057	−0.039	0.000	0.009
Seasonality ratio	0.640	0.187	0.380	0.654	0.893
Price variance	2.295	9.582	0.039	0.138	1.542
Autoregressive coefficient ϕ	−0.414	0.108	−0.541	−0.428	−0.258
Modified autoregressive coefficient $\phi\sigma^2$	−0.004	0.011	−0.007	−0.001	−0.0003
Industry size	6,429	8,409	603	3,534	14,178
Panel B: Retail industries ($N = 6$)					
Amplification ratio	0.855	0.606	0.291	0.782	1.947
Amplification difference	−0.018	0.028	−0.062	−0.004	0.006
Seasonality ratio	0.872	0.152	0.591	0.932	0.986
Price variance	0.161	0.12	0.039	0.164	0.364
Autoregressive coefficient ϕ	−0.436	0.125	−0.671	−0.403	−0.326
Modified autoregressive coefficient $\phi\sigma^2$	−0.0005	0.0003	−0.0009	−0.0004	−0.0002
Industry size	21,194	15,454	7,397	18,003	48,207
Panel C: Wholesale industries ($N = 18$)					
Amplification ratio	1.528	0.870	0.949	1.298	3.520
Amplification difference	0.003	0.008	−0.001	0.002	0.018
Seasonality ratio	0.610	0.105	0.460	0.624	0.761
Price variance	4.717	14.278	0.027	0.703	18.721
Autoregressive coefficient ϕ	−0.470	0.087	−0.553	−0.498	−0.289
Modified autoregressive coefficient $\phi\sigma^2$	−0.001	0.0004	−0.002	−0.0009	−0.0005
Industry size	9,786	6,262	3,448	9,522	22,645
Panel D: Manufacturing industries ($N = 50$)					
Amplification ratio	1.067	0.747	0.335	0.871	2.780
Amplification difference	−0.170	0.067	−0.040	−0.002	0.010
Seasonality ratio	0.623	0.196	0.320	0.663	0.883
Price variance	1.679	7.935	0.059	0.137	1.080
Autoregressive coefficient ϕ	−0.390	0.106	−0.504	−0.420	−0.239
Modified autoregressive coefficient $\phi\sigma^2$	−0.006	0.014	−0.010	−0.001	−0.0002
Industry size	3,449	5,241	453	1,967	8,020

preliminary tests of Hypotheses 1–3. Because of the small sample size of retail industries, these correlation results serve as our primary test of hypotheses for retail industries. Consistent with Hypothesis 1, we observe a significant negative association between the seasonality ratio and the amplification ratio in all panels. This association is strongest for retail industries (−0.87) and weakest for manufacturing industries (−0.29). Consistent with Hypothesis 2, we observe a significant positive association between price variance and the amplification ratio in the manufacturing industries. Regarding Hypothesis 3, we observe no significant correlation between the autoregressive coefficient ϕ and the amplification ratio. However, we observe a significant correlation between the amplification difference and the modified autoregressive coefficient $\phi\sigma^2$ in manufacturing industries.

Our primary tests of Hypotheses 1, 2, and 3 are presented in the regression results of Table 4. We show results of models using the amplification ratio and the amplification difference. Breusch-Pagan statistics indicate heteroscedastic residuals. Thus, all results are reported with robust standard errors. We noted no problems with multicollinearity. Outlier analysis using studentized residuals indicated several influential observations. Results are reported without outliers, but we note in the discussion below differences in results with outliers included.

The first and fourth columns of Table 4 present results of a regression model on the entire data set. We include indicator variables to control for different mean levels of amplification between industry sectors. The F statistics for the models are significant, and R^2s are greater than 0.50. We note no significant

Table 3 Spearman Correlation Coefficients of Study Variables

	Amplification ratio	Amplification difference	Seasonality ratio	Price variance	Autoregressive coefficient	Modified autoregressive coefficient	Industry size
Panel A: Entire data set ($N = 74$)							
Amplification ratio	1						
Amplification difference	0.96***	1					
Seasonality ratio	−0.42***	−0.42***	1				
Price variance	0.48***	0.46***	−0.25**	1			
Autoregressive coefficient ϕ	−0.07	−0.11	0.1	0.07	1		
Modified autoregressive coefficient $\phi\sigma^2$	0.27**	0.22	0.34***	0.16	0.29**	1	
Industry size	0.18	0.12	0.24**	0.12	−0.13	0.50***	1
Panel B: Retail industries ($N = 6$)							
Amplification ratio	1						
Amplification difference	0.94***	1					
Seasonality ratio	−0.87***	−0.94***	1				
Price variance	−0.14	−0.31	0.03	1			
Autoregressive coefficient ϕ	−0.54	−0.37	0.31	−0.09	1		
Modified autoregressive coefficient $\phi\sigma^2$	−0.26	−0.2	0.43	−0.54	−0.09	1	
Industry size	0.66	0.83**	−0.71	−0.66	−0.09	−0.14	1
Panel C: Wholesale industries ($N = 18$)							
Amplification ratio	1						
Amplification difference	0.91***	1					
Seasonality ratio	−0.75***	−0.54**	1				
Price variance	0.3	0.15	−0.48**	1			
Autoregressive coefficient ϕ	−0.04	0.03	−0.06	0.3	1		
Modified autoregressive coefficient $\phi\sigma^2$	−0.05	−0.23	−0.01	−0.02	0.29	1	
Industry size	−0.21	−0.24	0.23	0.02	0.22	0.37	1
Panel D: Manufacturing industries ($N = 50$)							
Amplification ratio	1						
Amplification difference	0.94***	1					
Seasonality ratio	−0.29**	−0.30***	1				
Price variance	0.54***	0.53***	−0.21	1			
Autoregressive coefficient ϕ	0.1	0.05	0.13	0.14	1		
Modified autoregressive coefficient $\phi\sigma^2$	0.40***	0.40***	0.34**	0.25	0.50**	1	
Industry size	0.05	0.003	0.28**	0.02	0.09	0.50***	1

Note. *** and **, significant at $p < 0.01$ and 0.05 levels, respectively.

differences in amplification ratios or amplification differences between levels of the supply chain after controlling for other covariates. This further emphasizes the notion that amplification does not increase from the lower to the higher levels of the supply chain. In support of Hypotheses 1 and 2, we report statistically significant coefficients for the seasonality ratio and price variability. The amplification ratio is negatively associated with the seasonality ratio and positively associated with price variability. We report no association between the amplification ratio and the autoregressive coefficient (Hypothesis 3). However, we note a positive significant coefficient between amplification difference and the modified autoregressive coefficient $\phi\sigma^2$. Industry size exhibits a positive and significant association with the amplification ratio but not the amplification difference.[24]

Columns 2, 3, 5, and 6 of Table 4 show regression models for the manufacturing and wholesale industries. The small population of retail industries (six industries) limits multivariate analysis of this sector. The R^2 values range from 0.32 in the manufacturing industries to 0.82 in the wholesale industries. The F statistics are significant in all models. Consistent with the overall sample and with Hypothesis 1, we

[24] With nonlogged data, industry size exhibits a positive and significant association with both amplification measures.

Table 4 Coefficients from Regression Models Predicting Amplification Ratios and Differences (t Values in Brackets)

	Dependent variable: Amplification ratio			Dependent variable: Amplification difference		
	All industries	Manufacturing	Wholesale	All industries	Manufacturing	Wholesale
Intercept	1.23***	1.20**	1.18	0.08**	0.09**	−0.01
	[3.09]	[2.66]	[0.57]	[2.55]	[2.29]	[−0.53]
Retail	0.09			0.01		
	[0.49]			[0.76]		
Wholesale	0.13			0.004		
	[0.88]			[0.73]		
Seasonality ratio	−1.63***	−1.49***	−5.68*	−0.09***	−0.09***	−0.05**
	[−3.92]	[−3.10]	[−1.87]	[−3.38]	[−2.90]	[−2.54]
Price variance	0.14***	0.16**	0.08*	0.003*	0.005	0.001
	[5.97]	[2.44]	[2.11]	[1.77]	[1.58]	[1.61]
Autoregressive coefficient ϕ	0.46	0.61	−1.56			
	[0.95]	[1.00]	[−0.77]			
Modified autoregressive coefficient $\phi\sigma^2$				2.72**	2.73**	−5.73
				[2.36]	[2.37]	[−1.06]
LN (Industry size)	0.12*	0.12	0.34	−0.004	−0.004	0.004
	[1.95]	[1.57]	[1.16]	[−1.06]	[−0.97]	[1.65]
R-squared	0.50	0.32	0.63	0.58	0.59	0.82
F statistic	29.79***	4.64***	18.46***	3.89***	3.77**	51.52***
N	70	48	17	71	48	17

Notes. ***, **, * significant at $p < 0.01$, 0.05, and 0.10 levels, respectively. Results reported with outliers deleted.

report a negative and significant coefficient for the seasonality ratio. In the manufacturing and wholesale subset and consistent with Hypothesis 2, we also report a positive and significant association between price variability and the amplification ratio but not the amplification difference. As in Column 1 of Table 4, we report no association between the amplification ratio and the autoregressive coefficient, but do report a significant positive association between the amplification difference and the modified autoregressive coefficient (Hypothesis 3). We report no significant results for industry size in these regressions.

The estimated seasonality ratio coefficients in Table 4 are quite large. For example, the coefficient estimated with the entire sample is −1.63. The seasonality ratio ranges from 0 to 1 so, for instance, an increase in the seasonality ratio of 0.25 would reduce the amplification ratio by 0.41. Given that the mean amplification ratio is 1.16 (Table 2), this reduction could switch an industry from being an amplifier to being a production smoother. A similar analysis demonstrates the other estimated coefficients for the seasonality ratio are also quite large.

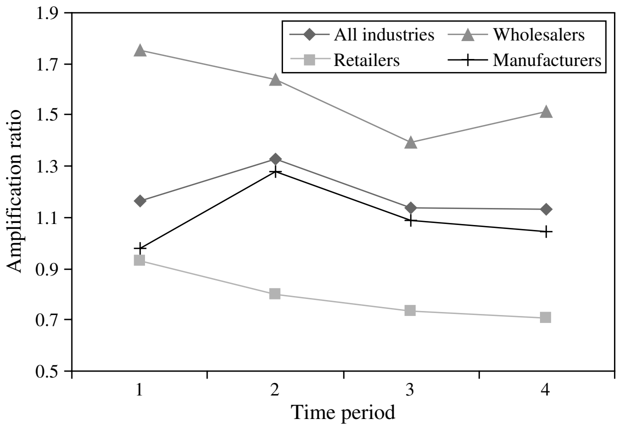
Results including outliers yielded one significant difference from those reported above. Two petroleum-related industries (one wholesale, one manufacturing) exhibited abnormally high price variance. The coefficient on price variance is not significant if these industries are included in the analysis.

Although we find a negative relationship between the amplification ratio and the seasonality ratio, both of those ratios have the variance of demand in the denominator. Thus, it is possible we are observing correlation between the two variables even though they are independent of each other. In the spirit of Gaur et al. (2005a), we estimate the following alternative specification to test if our results are induced by our specification of the variables:

$$\begin{aligned} &V[\mathit{Production}]\\ &= \phi_0 + \phi_1 V[\mathit{Demand}]\\ &\quad + \phi_2(V[\mathit{Demand}] - V[\mathit{Seasonally\ adjusted\ demand}])\\ &\quad + \phi_3[\mathit{Price\ index}] + \phi_4 V[\mathit{AR}(1)\ \mathit{coefficient}] + \phi_5 \mathit{Size}, \end{aligned}$$

where the variance of seasonally adjusted demand is evaluated as in the seasonality ratio. Results using

Figure 3 Amplification Ratios Estimated with a General Linear Model with Dummy Variables for Four Subperiods



1 = 1992–1995, 2 = 1996–1998, 3 = 1999–2001, 4 = 2002–2005

this alternative specification consistently show $\phi_2 < 0$, which provides a robustness test to our finding that the amplification ratio is negatively correlated with the seasonality ratio.

To explore if the amplification ratios are decreasing over time (Hypothesis 4), we divide our data series into four series (years 1992–1995, 1996–1998, 1999–2001, and 2002–2005). We estimate mean differences between amplification ratios in these quartiles using a general linear model. Figure 3 plots the mean amplification ratio by quartile for the entire sample, manufacturing, wholesale, and retail industries. While we do observe a decreasing trend for retailers and wholesalers, we observe no significant differences between the time series in any of the samples.

To further test for differences in time, we evaluate the variance of production, the variance of demand, and the two amplification measures for each industry (and the aggregate series) for each of the 14 years in the sample. We wish to determine if these series exhibit time trends. The Durbin-Watson d test indicates that most of these series have correlated errors, so we implement on each series generalized least squares (GLS) regression with AR(1) disturbances.[25] Table 5 reports the estimated time trend for the amplification measures (Columns 1 and 2) as well as the time trend for the variance of production (Column 3) and the variance of demand (Column 4). Fifty-one industries have decreasing amplification ratios (14 are significant), but only 23 have increasing amplification ratios (eight are significant). None of the aggregate series exhibit a time trend in their amplification ratios. Production variances display some tendency to decrease (43 negative slopes, 11 of them are significant), while demand variances display some tendency to increase (43 positive slopes, 15 of them are significant).

Each series contains only 14 observations, which raises potential concern with our t-statistic on the estimated slopes. Hence, we also tested for a significant slope coefficient with a Theil test, a nonparametric test (see Hollander and Wolfe 1999, pp. 200–201, for details). Results with the Theil test were qualitatively similar to those from the GLS estimation. For example, there are more industries with declining amplification ratios (48) than increasing amplification ratios (26) but most industries do not exhibit a significant trend (18 significant negative slopes, six significant positive slopes, and 50 insignificant slopes).

Some industries exhibit interesting patterns. Both the variance of demand and the variance of production is decreasing with general merchandisers, but because demand variance is decreasing faster than production variance, the amplification difference is actually increasing while the amplification ratio appears to be decreasing (though not significantly). We attribute this to the efforts of general merchandisers to extend the length of the fourth quarter holiday selling season, which reduces the variance of demand over time. Pharmaceutical manufacturing has experienced a strong increase in the amplification ratio (and we note the pharmaceutical wholesalers stongly exhibit the bullwhip effect), whereas

[25] Given a time series y with $t = 14$ observations, we first use ordinary least squares (OLS) to estimate

$$y_t = \beta_0 + \beta_1 t + u_t,$$

where u is assumed to be iid disturbance terms. We then use OLS to estimate

$$u_t = \rho u_{t-1} + v_t,$$

where v is assumed to be iid disturbance terms. We use the estimated autoregressive coefficient ρ to adjust the original series so that we can use ordinary least squares (OLS) to estimate

$$y_t^* = \beta_0^* + \beta_1^* t^* + u_t^*,$$

where $y_t^* = y_t - \rho y_{t-1}$, $\beta_0^* = \beta_0(1-\rho)$, $t^* = t - \rho(t-1)$, $u_t^* = u_t - \rho u_{t-1}$, and $\beta_1^* = \beta_1$. (OLS is now valid because u_t^* are now iid disturbance terms under the hypothesis that u_t follow an AR(1) process.)

Table 5 Trend in Amplification Measures, Production Variance, and Demand Variance (1992–2005)

	Estimated time trend coefficient			
	$V[Y]/V[D]$	$V[Y]-V[D]$	$V[Y]$	$V[D]$
Retail industries				
Aggregate retail series	0.0015	0.0002	−0.0001	−0.0003**
Building material and garden equipment and supplies dealers	−0.0054	−0.0001	0.0003**	0.0004**
Clothing and clothing accessory stores	−0.0244**	−0.0016	−0.0030**	−0.0014***
Food and beverage stores	−0.0053	0.0000	−0.0001	0.0000
Furniture, home furnishings, electronics, and appliance stores	−0.0221	−0.0008**	−0.0001	0.0007**
General merchandise stores	−0.0044	0.0024***	−0.0014***	−0.0038***
Motor vehicle and parts dealers	−0.0869	−0.0002	0.0002	0.0004*
Wholesale industries				
Aggregate wholesale series	0.0002	0.0000	0.0000	0.0000
Apparel, piece goods, and notions	−0.0096	−0.0004	−0.0019*	−0.0014***
Beer, wine, and distilled alcoholic beverages	−0.0064	−0.0002	0.0001	0.0003
Chemicals and allied products	−0.0122	−0.0001	−0.0001	0.0000
Drugs and druggists' sundries	−0.0623	−0.0003	−0.0003	0.0000
Electrical and electronic goods	0.0248	0.0001	0.0001	−0.0001
Farm product raw materials	0.1199*	0.0004	−0.0002	−0.0006
Furniture and home furnishings	−0.0596	−0.0004*	−0.0006*	−0.0002
Grocery and related products	−0.0353	−0.0001	−0.0002	−0.0001
Hardware, and plumbing and heating equipment and supplies	−0.0694*	−0.0004*	−0.0002	0.0002
Lumber and other construction materials	−0.0085	−0.0001	−0.0001	0.0001
Machinery, equipment, and supplies	0.0057	0.0000	0.0002	0.0003
Metals and minerals, for example, petroleum	−0.0023	−0.0005	−0.0008	−0.0001
Miscellaneous durable goods	0.0818*	0.0005***	0.0001	−0.0005*
Miscellaneous nondurable goods	−0.0118	−0.0002	−0.0005	−0.0005*
Motor vehicle and motor vehicle parts and supplies	−0.0612	−0.0002	0.0001	0.0003**
Paper and paper products	−0.0925	−0.0003	−0.0002	0.0001
Petroleum and petroleum products	0.0185	0.0001*	0.0002***	0.0001**
Professional and commercial equipment and supplies	−0.1103***	−0.0010***	−0.0003	0.0008*
Manufacturing industries				
Aggregate manufacturing series	−0.0119	−0.0001	−0.0001**	0.0000
Apparel	0.0157	0.0006***	−0.0004**	−0.0009***
Audio and video equipment manufacturing	0.0026	−0.0001	0.0004	0.0005
Automobile manufacturing	−0.0027	0.0000	−0.0018	−0.0018
Battery manufacturing	0.0966*	0.0039*	0.0047**	0.0006
Beverage manufacturing	0.2253	0.0009	0.0007	−0.0002
Communications equipment manufacturing, defense#	0.5394	−0.0025	0.0506***	0.0545*
Communications equipment manufacturing, nondefense#	0.0348**	0.0045	0.0003	−0.0043
Computer storage device manufacturing	−0.0323**	−0.0316***	−0.0112	0.0263*
Construction machinery manufacturing#	−0.0650***	−0.0015**	−0.0012***	0.0003
Dairy product manufacturing	−0.0451**	−0.0002	−0.0001	0.0000
Electric lighting equipment manufacturing	−0.0437	−0.0023**	−0.0004	0.0018***
Electrical equipment manufacturing#	−0.0532	0.0003	−0.0007	−0.0009
Electromedical, measuring, and control instrument manufacturing#	−0.0238	−0.0064***	0.0031***	0.0101***
Electronic computer manufacturing#	0.0256	0.0040**	−0.0010	−0.0047
Fabricated metal products#	0.0016	−0.0001	0.0000	0.0000
Farm machinery and equipment manufacturing	−0.0621***	−0.0014***	−0.0021***	−0.0007
Ferrous metal foundries#	−0.0229	−0.0006	−0.0019	−0.0011
Furniture and related products#	0.0021	−0.0001	−0.0001	−0.0001
Grain and oilseed milling	−0.0220	0.0004	0.0004	0.0001
Heavy duty truck manufacturing	−0.0524***	−0.0004*	0.0001	0.0004
Household appliance manufacturing#	−0.1278***	−0.0014***	−0.0011***	0.0001
Industrial machinery manufacturing#	−0.0240	−0.0079***	0.0006	0.0085***

Table 5 (cont'd.)

	Estimated time trend coefficient			
	$V[Y]/V[D]$	$V[Y]-V[D]$	$V[Y]$	$V[D]$
Iron and steel mills and ferroalloy and steel products manufacturing#	0.0536**	0.0001	0.0004	0.0000
Leather and allied products	0.0951	0.0014	0.0008	−0.0006**
Light truck and utility vehicle manufacturing	−0.0104	−0.0010	0.0000	0.0012
Material handling equipment manufacturing#	−0.0543**	−0.0031**	−0.0002	0.0028*
Meat, poultry, and seafood product processing	0.0231	0.0000	0.0001*	0.0001
Metalworking machinery manufacturing#	0.1117**	0.0025**	0.0008	−0.0016***
Mining, oil, and gas field machinery manufacturing#	0.2785	0.0055	0.0050	−0.0007
Miscellaneous products#	−0.0620	−0.0009*	−0.0007	0.0003
Nonmetallic mineral products	−0.0091	0.0000	−0.0001	0.0000
Other computer peripheral equipment manufacturing	−0.0270***	−0.0054	−0.0035*	0.0016
Other electronic component manufacturing#	−0.0780*	−0.0016***	0.0012**	0.0028***
Paint, coating, and adhesive manufacturing	−0.0065	−0.0001	0.0000	0.0001
Paperboard container manufacturing	−0.0485	0.0001	0.0002	0.0001
Pesticide, fertilizer, and other agricultural chemical manufacturing	−0.0035	0.0001	−0.0007	−0.0003
Petroleum and coal products	−0.0150	0.0010***	0.0015***	0.0004***
Pharmaceutical and medicine manufacturing	0.5210***	0.0037***	0.0036**	−0.0001
Photographic equipment manufacturing#	−0.2682**	−0.0025	−0.0011	0.0015
Plastics and rubber products	0.0494	0.0002*	0.0001	−0.0002**
Printing	−0.0694	−0.0002	−0.0001	0.0000
Pulp, paper, and paperboard mills	−0.0357	−0.0001	0.0000	0.0001
Search and navigation equipment manufacturing, defense#	−0.0088	−0.0264	0.0000	0.0275
Search and navigation equipment manufacturing, nondefense#	−0.0642	−0.0009	−0.0060	−0.0088
Textile products	0.0126	0.0002	−0.0003	−0.0004*
Textiles	0.0329***	0.0007**	−0.0002	−0.0008*
Tobacco manufacturing	−3.3627	−0.0233***	−0.0306***	0.0021
Transportation equipment#	−0.0045	−0.0013	−0.0004	0.0009
Ventilation, heating, air conditioning, and refrigeration#	−0.0877***	−0.0020**	−0.0006***	0.0025**
Wood products	−0.0092	−0.0001	0.0000	0.0001

Notes. $V[Y]$ = Variance of production; $V[D]$ = Variance of demand.
Industry for which demand and sales data are available. In all other industries, sales is used as a proxy for demand.
***, **, * significant at $p < 0.01$, 0.05, and 0.10 levels, respectively.

household appliance and photographic equipment manufacturing have experienced sharp declines in amplification ratios.

Overall, as with the results from the general linear model, there is some evidence that amplification ratios are decreasing, but the evidence is not overwhelming. Therefore, our results do not appear to be driven by any particular subperiod of our sample.

6.3. Summary of Results

We conclude that we do not observe the bullwhip effect among retailers, and we generally do not observe it among manufacturers. Although the majority of wholesalers amplify, there is little evidence that demand volatility is highest among manufacturers and least among retailers.

We find strong support that seasonality influences the amplification ratio (Hypothesis 1): Industries with seasonality tend to smooth production relative to demand, whereas industries without seasonality tend to amplify. Our results for seasonality apply for the entire sample as well as for each subsample representing the three main levels of the supply chain. We also find some support that price variability, which is a proxy for promotion activity as well as cost shocks, contributes to amplification (Hypothesis 2). Most industries have negatively correlated demand shocks and the amount of demand correlation does not significantly influence the amplification ratio (Hypothesis 3), but we do find support that it increases the amplification difference. Finally, we find modest evidence that amplification measures are decreasing, but the overall pattern is mostly stable over our sample period (Hypothesis 4).

7. Discussion

Previous studies find the bullwhip effect in numerous examples from individual products and most U.S. industries (i.e., the variance of production exceeds the variance of sales). In contrast, we find that most retailers are production smoothers, as are the majority of manufacturing industries. Only wholesalers appear to consistently amplify in our sample. However, even though wholesalers amplify, manufacturing demand is not more volatile than demand at the lower levels in the supply chain. There is even some evidence that retail demand is the most volatile, which is inconsistent with the bullwhip effect.

Our results do appear to contrast with the existing literature: The economics literature finds production to be more volatile than sales in the vast majority of industries, and the operations management literature offers numerous examples of the bullwhip effect. However, we observe that most wholesale and many manufacturing industries exhibit the bullwhip effect and find that price volatility contributes to the bullwhip effect, as predicted by theory. The key difference with our analysis is that we work with seasonally unadjusted data. We conjecture that seasonality provides a strong motivation for firms to attenuate variability and this mitigates the incentives to amplify. It follows that the industries with the least amount of seasonality should also be the ones that amplify the most, and we find strong support for that hypothesis. Furthermore, our results with seasonally adjusted data are nearly identical to previous findings; that is, when a key reason to smooth demand is removed, then nearly all industries exhibit amplification behavior.

It is possible that our results differ from the existing literature because the time period of our sample does not overlap with the time period of most previous studies. For example, it is possible that before 1992 the bullwhip effect was more prominent in the U.S. economy. Although we find some evidence that amplification ratios decreased over our sample period, we feel that it is unlikely there was a substantial decrease in ratios that occurred just before our sample. Unfortunately, it is difficult to test this conjecture given the lack of seasonally unadjusted data before our sample.[26]

Related to the issue of time, it is also possible that special circumstances in the U.S. economy during our sample period contributed to our findings. The U.S. economy grew significantly during the 1990s, so the lack of a bullwhip effect may be due to firms experiencing capacity constraints: Retailers and manufacturers may de facto smooth production because peak demand exceeds their capacity. We do not think our evidence points to this explanation. The U.S. economy took a downturn in the last quartile of our data, so if excess capacity enables the bullwhip effect, the fourth quartile of our sample should have higher amplification ratios than the third, but we do not observe such a pattern.

Clearly, the issue of seasonality is central to our findings. We admit that it is debatable whether the amplification measures should be taken with unadjusted data (as we do) or with seasonally adjusted data. We are comfortable working with unadjusted data primarily because even though seasonality is quite inconvenient when working with theoretical models (most of the modeling work on the bullwhip effect assumes stationary demand), firms must respond to actual demand, not seasonally adjusted demand. In other words, production costs are likely to depend on the variability of production, not just the variability of seasonally adjusted production.

7.1. Limitations and Extensions

Our analysis is conducted on industry-level data, which have strengths and limitations. Industry-level data are available for a wide range of industries and different levels of the supply chain, which allows us to study supply chain amplification across a broad spectrum of the U.S. economy. Previous studies of the bullwhip effect have similarly studied industry-level

[26] There were two other theories to explain amplification in the U.S. industry data. Fair (1989) suggests that it is due to nonphysical unit measures. The use of dollar-denominated units in our data is an issue, but it does not appear to prevent us from finding production smoothing. Others argued that the bullwhip effect could be present in the United States but possibly not in Japan. We are unable to comment on whether amplification ratios are greater or smaller between the United States and Japan, but it appears that production smoothing also occurs in the United States.

data, but even without a connection to past research, industry-level data are interesting on their own. Most firms sell many products to many other firms in a downstream industry. With respect to procurement, labor, and capacity planning, a firm should be concerned with the total volatility received by all its customers in addition to the volatility received from any single product. In other words, industry-level volatility is relevant to some (but not all) operational decisions. However, it is not possible to conclude from industry-level volatility whether amplification occurs at the firm, division, category, or product level. For example, it is possible that firms exhibit the bullwhip effect but the industry does not. The opposite is also possible; firms could production smooth but the industry exhibits the bullwhip effect. Whether aggregation preserves or masks the bullwhip effect or production smoothing depends on the correlation of production and demand across the units being aggregate (firms, products, etc.) and on the particular causes of amplification in place.

It should be noted that aggregation does not necessarily cause a bias. For example, Caplin (1985) shows that if the bullwhip effect is due to (S, s) policies, then it is preserved under aggregation no matter what the demand correlation structure is. Using quarterly data (as opposed to our monthly data), Allen (1997) finds that aggregation tends to preserve whether the amplification ratio falls above or below one; he also finds that aggregation moves the ratios closer to one (i.e., they increase up toward one or decrease down toward one). Thus, even though we believe our results based on industry-level data are informative, further research is needed on the prevalence of the bullwhip effect at finer levels of aggregations.

In addition to product aggregation, there are issues with time aggregation. We work with monthly data, which is the appropriate time interval for some decisions such as labor planning in a warehouse, but it is not the appropriate time interval for other decisions. For example, daily or weekly time intervals are relevant for some transportation decisions, and quarterly and yearly data are more relevant for significant capacity adjustments (such as building a new plant). Shorter time intervals will have different seasonality patterns than our monthly patterns, and those patterns will influence the degree of amplification. For longer time intervals, it may be necessary to keep the long-run trend in the data series. (Recall that we removed long-run trends by first differencing.) Hence, different levels of time aggregation are likely to require different empirical methods. Furthermore, just as with product aggregation, it is not possible to conclude how the degree of amplification will change from one time interval to another.

With our data, as we have already mentioned, it is difficult to make comparisons across different levels of the supply chain because it is not possible to construct linear supply chains. It would be worthwhile to study a data set of products or firms that can be more explicitly linked into a supply web, to determine how the pattern of amplification changes at different positions in the network. On a related point, we do not have first-hand knowledge of the methods used by Census to collect these data, nor of the accuracy of the data. For example, Census surveys a limited set of firms and the sample of firms changes over time, so Census performs adjustments before reporting the data. As a result, it would be useful to collect and analyze a primary data set. Such an effort ideally would collect data on actual orders in addition to physical inflow and inventory data.

Our analysis investigates whether the incentives to amplify demand dominate the incentives to attenuate demand. As already discussed, our analysis does not explicitly test the theories on the causes of the bullwhip effect, for example, whether temporary cost shocks lead to amplication or whether restricted capacity leads to shortage gaming. Such tests would surely be useful but require different data. In particular, detailed data are needed on a specific cause of the bullwhip effect, such as measures of fixed ordering costs.

8. Conclusion

It is clear that the issue of supply chain volatility has received considerable attention both in economics and in operations management (although not at the same time). Economists have focused on explaining why they did not observe production smoothing, but we suggest that production smoothing is indeed relatively common, especially when conditions are most favorable to do so (when there is a substantial amount of seasonality). In operations management, the focus has been on identifying causes of the bullwhip effect and suggesting mitigation strategies based on those causes. Our findings are generally consistent with those causes, but because there are strong forces that mitigate the bullwhip effect (again, seasonality), the bullwhip effect is often not observed in industry-level data. This is indeed good news for firms and their suppliers. Now, attention should turn toward probing data from individual firms and individual products so that we can deepen our understanding of this phenomenon.

Acknowledgments
The authors thank seminar participants at the Fuqua School of Business, Harvard Business School, INSEAD, Kenan-Flagler Business School, LeBow College of Business, Olin School of Business, Sloan School of Management, Penn State Smeal College of Business, Stanford Graduate School of Business, Tuck School of Business (Supply Chain Thought Leaders Conference), Sauder School of Business, University of Texas at Dallas, University of Washington, the Wharton School of Business, and The Workshop on Empirical Research in Operations Management held at the Wharton School of Business for their helpful comments. The authors also benefited from helpful discussions with Alan Blinder, Charlie Fine, Hau Lee, Serguei Netessine, Karl Ulrich, John Sterman, and Yu-Sheng Zheng. The first author thanks the Fishman-Davidson Center of the University of Pennsylvania for financial support.

References

Abel, A. 1985. Inventories, stock-outs and production smoothing. *Rev. Econom. Stud.* **52** 283–293.

Allen, D. 1997. Do inventories moderate fluctuations in output? *Federal Reserve Bank St. Louis Rev.* (Jul.-Aug.) 39–50.

Anderson, E., C. Fine, G. Parker. 2000. Upstream volatility in the supply chain: The machine tool industry as a case study. *Production Oper. Management* **9** 239–261.

Aviv, Y. 2001. The effect of collaborative forecasting on supply chain performance. *Management Sci.* **47** 1326–1343.

Aviv, Y. 2002. Gaining benefits from joint forecasting and replenishment processes. *Manufacturing Service Oper. Management* **4** 55–74.

Aviv, Y. 2003. A time-series framework for supply chain inventory management. *Oper. Res.* **51** 210–227.

Baganha, M., M. Cohen. 1998. The stabilizing effect of distribution center inventory. *Oper. Res.* **46** 72–83.

Beason, R. 1993. Tests of production smoothing in selected Japanese industries. *J. Monetary Econom.* **31** 381–394.

Blanchard, O. 1983. The production and inventory behavior of the American automobile industry. *J. Political Econom.* **91** 365–400.

Blanchard, O., J. Simon. 2001. The long and large decline in U.S. output volatility. *Brookings Papers Econom. Activity* **1** 1135–1174.

Blinder, A. 1981. Retail inventory investment and business fluctuation. *Brookings Papers Econom. Activity* **2** 443–505.

Blinder, A. 1986. Can the production smoothing model of inventory behavior be saved? *Quart. J. Econom.* **101** 431–453.

Blinder, A., L. Maccini. 1991. Taking stock: A critical assessment of recent research on inventories. *J. Econom. Perspectives* **5** 73–96.

Boettcher, J. C., L. M. Gaines. 2004. *Industry Research Using the Economic Census*. Greenwood Press, Westport, CT.

Brown, M., A. Forsythe. 1974. Robust tests for the equality of variances. *J. Amer. Statist. Assoc.* **69** 364–367.

Cachon, G. 1999. Managing supply chain demand variability with scheduled ordering policies. *Management Sci.* **45** 843–856.

Cachon, G., M. Lariviere. 1999. Capacity choice and allocation: Strategic behavior and supply chain performance. *Management Sci.* **45** 1091–1108.

Caplin, A. S. 1985. The variability of aggregate demand with (s, s) inventory policies. *Econometrica* **53** 1395–1409.

Chen, F., Z. Drezner, J. Ryan, D. Simchi-Levi. 2000. Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information. *Management Sci.* **46** 436–443.

Eichenbaum, M. 1989. Some empirical evidence on the production level and production cost-smoothing models of inventory management. *Amer. Econom. Rev.* **79** 853–864.

Fair, R. 1989. The production smoothing model is alive and well. Working Paper 2877, NBER Working Paper Series.

Gaur, V., M. Fisher, A. Raman. 2005a. An econometric analysis of inventory turnover performance in retail stores. *Management Sci.* **51** 181–193.

Gaur, V., A. Giloni, S. Seshadri. 2005b. Information sharing in a supply chain under ARMA demand. *Management Sci.* Forthcoming.

Ghali, M. 1987. Seasonality, aggregation and the testing of the production-smoothing hypothesis. *Amer. Econom. Rev.* **77** 464–469.

Gilbert, K. 2005. An ARIMA supply chain model. *Management Sci.* **51** 305–310.

Graves, S. 1999. A single item inventory model for a nonstationary demand process. *Manufacturing Service Oper. Management* **1** 50–61.

Hamilton, J. 1994. *Times Series Analysis*. Princeton University Press, Princeton, NJ.

Hammond, J. 1994. Barilla SpA (a). HBS Case 9-694-046, Harvard Business School, Boston, MA.

Hollander, M., D. Wolfe. 1999. *Nonparametric Statistical Methods*, 2nd ed. Wiley-Interscience, Hoboken, NJ.

Holt, C. C., F. Modigliani, J. P. Shelton. 1968. The transmission of demand fluctuations through distribution and production systems: The tv-set industry. *Canadian J. Econom.* **14** 718–739.

Kahn, J. 1987. Inventories and the volatility of production. *Amer. Econom. Rev.* **78** 667–679.

Kahn, J. 1992. Why is production more volatile than sales? Theory and evidence on the stockout-avoidance motive for inventory-holding. *Quart. J. Econom.* **2** 481–510.

Krane, S., S. Braun. 1991. Production smoothing evidence from physical product data. *J. Political Econom.* **99** 558–581.

Lee, H., V. Padmanabhan, S. Whang. 1997a. Information distortion in a supply chain: The bullwhip effect. *Management Sci.* **43** 546–558.

Lee, H., V. Padmanabhan, S. Whang. 1997b. The bullwhip effect in supply chains. *MIT Sloan Management Rev.* **38** 93–102.

Lee, H., V. Padmanabhan, S. Whang. 2004. Comments on "Information distortion in a supply chain: The bullwhip effect." *Management Sci.* **50** 1887–1893.

McConnell, M., G. Perez-Quiros. 2000. Output fluctuations in the United States: What has changed since the early 1980s? *Amer. Econom. Rev.* **90**(5) 1464–1476.

Miron, J., S. Zeldes. 1988. Seasonality, cost shocks, and the production smoothing model of inventories. *Econometrica* **56** 877–908.

Mollick, A. V. 2004. Production smoothing in the Japanese vehicle industry. *Internat. J. Production Econom.* **91** 63–74.

Mosser, P. 1991. Trade inventories and (S,s). *Quart. J. Econom.* **106** 1267–1286.

Raghunathan, S. 2001. Information sharing in a supply chain: A note on its value when demand is nonstationary. *Management Sci.* **47**(4) 605–610.

Ramey, V. 1991. Nonconvex costs and the behavior of inventories. *J. Political Econom.* **99** 306–334.

Rossana, R. 1998. Structural instability and the production-smoothing model of inventories. *J. Bus. Econom. Statist.* **16** 206–215.

Sobel, M. 1969. Production smoothing with stochastic demand *i*: Finite horizon case. *Management Sci.* **16** 195–207.

Sterman, J. 1989. Modeling managerial behavior: Misperceptions of feedback in a dynamic decision-making experiment. *Management Sci.* **35** 321–339.

Terwiesch, C., J. Ren, T. H. Ho, M. Cohen. 2005. Forecast sharing in the semicondutor equipment supply chain. *Management Sci.* **51** 208–220.

U.S. Bureau of Economic Analysis. 2006. National income and product accounts (NIPA), tables 2AUI and 2BUI. Implicit Price Deflators for Manufacturing and Trade Sales, www.bea.doc.gov/bea/dn/nipaweb/nipa_underlying/SelectTable.asp?Benchmark=P.

U.S. Census Bureau. 2005. 1997 Census. http://www.census.gov/epcd/ec97/us/US000.HTM.

U.S. Census Bureau. 2006a. Current business reports, series BR/04-A. Annual Benchmark Report for Retail Trade and Food Services: January 1992 through February 2006, Washington, D.C. Available at http://www.census.gov/prod/www/abs/br_month.html.

U.S. Census Bureau. 2006b. Current business reports, series BW/04-A. Annual Benchmark Report for Wholesale Trade: January 1992 through January 2006, Washington, D.C. Available at http://www.census.gov/prod/www/abs/bw_month.html.

U.S. Census Bureau. 2006c. Manufacturers' shipments, inventories, and orders: 1992–2002, M3-1(02). U.S. Government Printing Office, Washington, D.C. Post-2002 data were obtained at http://www.census.gov/indicator/www/m3/hist/naicshist.htm.

West, K. 1986. A variance bounds test of the linear quadratic inventory model. *J. Political Econom.* **94** 374–401.