# Bullwhip in a Spanish Shop

Richard Lai

Harvard Business School

18. September 2005

# Negotiation, Organizations and Markets Research Papers

Harvard NOM Research Paper No. 06-06


Bullwhip in a Spanish Shop


**Richard Lai**

# Bullwhip in a Spanish Shop

## 1. Introduction

The bullwhip effect describes how demand is increasingly distorted as it gets passed up a supply chain. It is one of the central observations in operations management (e.g., Lee, et al. (2004)) and economics (e.g., Blinder and Maccini (1991)). There are many theories of what cause it, but not enough facts. Some theories argue that the bullwhip is caused by rational factors such as information distortion (e.g., Lee, et al. (1997b), Chen, et al. (2000)). Others suggest behavioral causes such as managers' underestimating the amount already ordered but not delivered (e.g., Sterman (1989), Croson and Donohue (forthcoming)). Who is right?

To answer this question, I use a novel dataset on 3,745 SKUs from Sebastian de la Fuente, the sixth largest supermarket chain in the Basque region of Spain. The 108,605 observations have been collected for almost 2 ½ years, from January 1990 through May 1992 (please see Aguirregabiria (1999); (2004)).

I find that the most prominent cause is the retailer's batching of orders to suppliers. Behavioral causes also seem to matter. I cannot find evidence for other rational causes proposed, such as cost shocks, demand correlation, or gaming in anticipation of shortages. While the findings are based on a large heterogeneous sample of SKUs from a modern supermarket, it is still data from one supermarket, so I advocate care in out-of-sample inferences. Despite this, I believe the paper still makes a contribution, by reporting the first tests of the bullwhip effect at the SKU level.

This report could also be useful in practice. It is crucial for managers to pinpoint causes, which imply very different solutions. For example, if the bullwhip were caused by gaming in anticipation of shortages, information sharing might be less useful a solution than incentive re-design. If behavioral causes are important, then the solution might involve sensitizing managers to their biases.

Poincare, when asked why econometricians assume normality of disturbances, is said to have replied: "The experimenters think it is a mathematical theorem, the mathematicians think it is an empirical fact." With little empirics so far, the bullwhip effect has the potential to take on a life of its own, too. With this paper, I hope it can be put on more solid ground.

## 2. The Bullwhip Effect

Clark (1917) is the first to start a lively discussion among economists when he points out that capital formation is amplified as it progresses up a supply chain. Metzler (1941) is credited by Blinder and Maccini (1991) to have started the same discussion, but with a focus on inventory. In the operations management context, Lee, et al. (1997b) have done the most to expound on the subject. Because so much has been written on the bullwhip effect, it is helpful to be clear about what it is and what I mean it to be in this paper.

There are at least five dimensions to the effect, which I label as construct, statistic, scope, party, and aggregation. By *construct*, I mean what is the thing that is being amplified: goods (e.g., deliveries from suppliers to customers) or information (e.g., orders from customers to suppliers). By *statistic*, I mean whether the amplification is a ratio of variances, maxima, or say, inter-quartile ranges. By *scope*, I mean whether the reference is to gross amplification, or amplification net of predictable factors such as seasonality. By *party*, I mean whether the setting is just one firm in a supply chain or all firms in the chain. Finally, *aggregation* is the level of analysis, such as at the industry, firm, category, or SKU levels.

In this paper, I focus on the flow of goods, not information. Goods flows have the advantage of being a real phenomenon that affects costs. Information flows such as orders account for quantities such as stock outs and back orders, and thus might be more useful normatively. In this paper, I choose goods over information because of data availability and because the subject of this paper is on positive description of real effects. I should also mention that models in economics (e.g., Kahn (1992), pg. 483; see also Lee, et al. (1997b), pg. 1877) have predictions of the bullwhip effect that are invariant to assumptions about back orders, so focusing on goods may not be a big departure from focusing on information orders.

I use *both* maxima and variances for the statistic dimension. Maximum quantities are indicative of costs of disruption or capacity built up. They are also probably what practitioners remember the most when they comment on the existence of the bullwhip effect. Variances, however, are a better measure of volatility, and have practical implications for allocation of production to capacities (e.g., Fisher and Raman (1996)).

I report results of *both* gross amplification and amplification net of seasonality. Gorman and Brannon (2000) and Cachon, et al. (2005) argue that the former is a better measure because it concerns real quantities that firms manage. I follow the literature in economics that partitions seasonality and non-seasonal components so as to gain better insight into both (e.g., Holt, et al. (1968), Miron and Zeldes (1988)). It is also be possible to *define* the bullwhip effect as only the amplification caused by the four factors described in Lee, et al. (1997b). However, I choose not to use this restriction here, since I am testing for other possible causes of amplification.

On the "party" dimension, I consider amplification at one party in the chain, so one way to qualify this paper is that it is about the contribution by a retailer to the bullwhip effect along the supply chain. I readily acknowledge that the bullwhip effect might be alternatively defined as a phenomenon on the entire supply chain, so that it requires either a strong form (*every* party in the supply chain shows amplification) or weak one (the upper-most party shows bigger variation than the lower-most one). Any study of this latter phenomenon is faced with the formidable empirical challenge of tracing SKUs through the supply chain. I am not aware of any work that manages to do this.

Finally, in terms of aggregation, I focus on SKUs, categories, and category groups. There is already a large literature that focuses on different industries – e.g.: Mack (1953) looks at the shoe-leather-hide industry, Holt, et al. (1968) at TV sets, Blanchard (1983) at automobiles, Miron and Zeldes (1988) at food, tobacco, apparel, chemicals, petroleum, and rubber, Mosser (1991) at 9 retail industries, Hammond (1994) at pasta, Lee, et al. (1997a) at soup, Anderson, et al. (2000) at machine tools, Terwiesch, et al. (2005) at semiconductors. There are also those that look at sets of industries, such as Gorman and Brannon (2000) at 14 manufacturing industries in the U.S. economy, Krane and Braun

(1991) at 28, and Blinder (1986), Bivin (1996) and Cachon, et al. (2005) at all industries in the U.S. economy, with the latter making the crucial contribution of studying industries along a supply chain. The gap is in large-sample investigation of the bullwhip at lower levels of analysis (see Cachon, et al. (2004), Cachon, et al. (2005), Lee, et al. (2004)). I should mention, however, that there is some validation at finer levels that come from experiments, such as the work by Croson and Donohue (2003), Carlos (2004), and Steckel, et al. (2004), and by simulation, such as those by Ge, et al. (2004) and Chatfield, et al. (2004) (a recent survey is in Kleijnen and Smits (2003)).

However the bullwhip effect is defined, there is the question of its existence. The literature is ambiguous. Much of the economics literature has documented its existence (e.g., Blinder (1986)), but Gorman and Brannon (2000) argue that these are due to their not adjusting for seasonality. Using seasonality-adjusted data, they find that 10 out of 14 manufacturing industries in the U.S. economy do not exhibit amplification. Cachon, et al. (2005) argue the opposite, that it is seasonal adjustments that bias studies toward finding the bullwhip effect. Using unadjusted data, they report mixed evidence of the bullwhip effect: 84% of wholesale industries, 38% of manufacturing ones, and just 14% of retail industries. This is yet again different than Bivin (1996), who also uses unadjusted (but detrended) data and find that only one of the 24 U.S. industries exhibit strictly no amplification. Miron and Zeldes (1988), using both adjusted and unadjusted data, report amplification ratios as high as 1.95 for tobacco and as low as 0.50 for apparel.

No matter how pervasive the bullwhip might be at the industry level, an intriguing unanswered question remains: does it exist at the SKU or category levels? This question is important since it is at these finer levels that define much of operations management. The answer might also shed light on the industry-level analyses, since we can now see if the lack of evidence for the bullwhip at those higher levels might be due to aggregation from SKU to category to industry. Caplin (1985) shows that the bullwhip effect is aggregation-neutral with an ($S$ $s$) policy under certain conditions. However, Caplin's argument does not apply in a model using the certainty-equivalence linear decision rules derived from quadratic cost functions first proposed by Holt, et al. (1960). Further, for the retail industry, Summers (1981) argues that the (*S s*) policy is not even a realistic description. Finally, Caplin's argument is about aggregation of the *same* SKU from multiple retailers, but in contrast, macroeconomic data aggregates over a heterogeneous set of SKU's from many retailers. Such cross-SKU aggregation might still cause serious biases in estimation.

To sum up, while this paper concentrates on identifying causes of the bullwhip effect, it also documents how pervasive the bullwhip is at the SKU level and whether the effect aggregates up to category or category group levels. Given the dearth of empirics, evidence on pervasiveness and aggregation could itself be a contribution to the literature.

## 3. Causes

If the existence of the bullwhip effect is still empirically under-studied, its causes are even less so. I enumerate the possible causes as identified in the literature. The list is not meant to be mutually exclusive or collectively exhaustive.

I group the possible causes into two broad classes: rational and behavioral. Rational causes can be justified with decision rules. They can be set within the context of one person or firm's optimization, or in a collaborative or non-cooperative game. Behavioral causes are those that are associated with systematic biases in human faculties. The utility of this rational vs. behavioral classification is practical. For example, Croson, et al. (2004) make the point that one can think of addressing rational causes as an exercise in aiming for the theoretical optimal (but see Conclusion for arguments that eliminating the bullwhip may not be financially optimal),, and understanding behavioral causes might be thought as aiming for the practically possible. Fixing rational problems might involve redesigning incentives, systems, or structure, while fixing behavioral ones might include a disproportionate focus on the human element: retraining, socializing, perhaps even reducing it through automation. Nevertheless, the danger in this sort of rational vs. behavioral classification is that one could get bogged down in a debate about whether a cause is of one type or the other. I risk proposing this classification with the hope that the utility of the classification outweighs this risk. The utility is in conceptually simplifying the causes, and possibly their implied solutions, through grouping. The risk is that the grouping is unsound. I leave it to the reader to make her own judgment on the matter, and push on to the focus this paper on identifying the existence of individual possible causes and, to the extent possible, their explanatory power for the bullwhip effect. I should also mention that there are other classification schemes. For instance, Daganzo (2004) argues that policy-oriented causes are better explanations of the bullwhip effect than demand-oriented ones. In the interest of parsimony, I will not classify the causes further.

There are four causes identified in the literature that could be classified as rational. First, the bullwhip could arise from cost shocks. This idea is generally associated with economists such as Blinder (1986), who propose it to explain the empirical volatility of production relative to sales. In a retail setting, Lee, et al. (1997b) identify this as price variation. The intuition is that variations such as trade promotions induce retailers to order more to take advantage of low costs. Empirically, this does not seem to be a consensus. For example, Miron and Zeldes (1988) show evidence that rules out cost shocks as a potential cause, at the industry level.

The second possible cause of the bullwhip effect comes from a setting with three conditions: correlated demand, stock out costs, and time lags in replenishing supplies (see Lee, et al. (1997b), Chen, et al. (2000)). The story here is that at every stage of a multi-echelon system, each party "plays safe" by ordering a bit more. Lee, et al. (1997b) provide a proof based on an AR(1) (autoregressive of order one) demand model with lags. Several authors use alternative demand models and obtain the same bullwhip result. For example, Graves (1999) uses an integrated moving average of order (0,1,1) process and Gilbert (2005) uses an ARIMA process. Chatfield, et al. (2004) use simulation to verify some of the models. Dejonckheere, et al. (2003) show that such amplification is robust to any forecasting method, if the firm uses an order-up-to replenishment policy.

A stronger variant of the above story is proposed by Kahn (1987). Specifically, he proposes that the bullwhip can obtain without replenishment lag, and with just the other two conditions: AR(1) demand and stock out costs. The approach is to have a linear-quadratic cost model; if the target level of inventory is proportional to expected next-period sales, then it can be shown theoretically that the bullwhip effect obtains. Empirically, this variant is usually pitted against theories based on cost shocks (the first cause), and the debate continues mostly because the studies are not comparable. For example, West (1990) looks at volatilities of inventories versus GNP and concludes that cost shocks are more important, while Kahn (1992) looks at the automobile industry and concludes the reverse, although he acknowledges the evidence is "circumstantial."

The third possible cause is that retailers batch their orders, resulting in correlated demand for suppliers. This idea is first elaborated by Lee, et al. (1997b) and further developed by Cachon (1999). Cetinkaya and Lee (2000) describe a variant under a VMI (vendor-managed inventory) regime, based on batching by the supplier rather than by retailers, that also causes the bullwhip effect. However, Baganha and Cohen (1998) show that batching might be counter-balanced by negative correlation among the retailers, thus reducing variance at the supplier. Whether batching is important might be resolved empirically.

The fourth possible cause is also identified by Lee, et al. (1997b): gaming by retailers in the face of supply shortages. Cachon and Lariviere (1999a) work out the game between one supplier and multiple downstream retailers to show that truth-telling is not a best response for either supplier or retailers, and the supplier would not increase capacity while retailers would over-order. In Cachon and Lariviere (1999b), they show that a more innovative way for the supplier to allocate capacity, the turn-and-earn regime common in the automobile industry, is still extractive if capacity is tight. In short, theory predicts that gaming can cause the bullwhip.

The above rational causes can be nicely stacked against a behavioral view, in which causes are due to systematic biases in cognition.

The first behavioral possibility is perceptual bias. Forrester (1958), Sterman (1989), Croson and Donohue (forthcoming) propose that managers ignore factors like lag times and goods ordered but not yet arrived. This argument is in line with the idea of under-reaction in psychological biases (e.g., Tversky and Kahneman (1973)). Another bias stems from representativeness, or over-reaction to recent signals. Along this line, theory predicts that the most recent order size dominates the decision on how much to order in the next period. It is unclear, however, whether this leads to amplification or attenuation. Representativeness amplifies if the underlying volatility is lower, and attenuates otherwise.

The second possible cause in the behavioral category has a sociological flavor. Croson, et al. (2004), in a clever experiment, document that retailer-subjects worry enough about the poor execution ability of their suppliers that they over-order, creating "coordination stocks" as buffer. It might be argued that this is a rational argument, as amenable to modeling as others. Without passing judgment on this, I list this as behavioral as classified by its proponents.

## 4. Data

During the period of the dataset (1990-92), Sebastian de la Fuente has been held substantially by Banco Bilbao Vizcaya (BBV), Spain's second biggest bank, through the latter's Corporacion Financiera de Distribucion (Cofidisa). Before 1989, Sebastian was held by Tabaclera, a state-owned company. As of 1989, the company has a turnover of about 116 billion ptas (US$983 million) and about 70 outlets, mainly located in Vizcaya, Guipuzcoa and Cantabria. Logistically, Sebastian operates a distribution center that intermediates between its outlets and suppliers. There are, however, some items that bypass the center and are sent direct from suppliers to outlets daily. These include some perishable items like fish and others whose suppliers have efficient systems. Other than these, the logistics cycle is monthly. For goods going through the distribution center, the lag is one month. To the extent that this is unrepresentative of other retailers, the conclusions here need to be modified.

My dataset consists of monthly information at the distribution center, including prices, markups, sales delivered to outlets, supplies received from suppliers, inventory (at the center, not at the outlets), and auxiliary information such as whether an SKU is on trade promotion, whether it is a private label, or whether it is stocked out. The SKUs are grouped into categories, and the categories into category groups (see summary statistics below). The company has on record 8,742 SKUs during the period, many of which are of no or low volumes. My dataset has 3,745 of the higher volume items, but even in this set, there are 1,363 SKUs with just one month of sales. For parts of my analysis, I create a sub-sample of all 534 SKUs that have the full 29 months of sales and whose logistics pass through the distribution center. Therefore, within this "534-subsample," ordering and inventory decisions are not confounded by decisions not to carry SKUs altogether.

The dataset made available to me is superior in several ways, apart from being at the SKU level and has rich detail. First, quantities are measured in physical units, rather than proxied using dollar amounts as in previous studies (e.g., Fair (1989), Krane and Braun (1991), Humphreys (2001)). This avoids measurement and accounting problems associated with dollar costing inputs and outputs when there is inventory involved. This does not mean that physical unit measures are superior on all counts. For instance, one can legitimately argue that firms should be more concerned with dollar values than physical units. But when dealing with operational parameters such as the volatility of SKUs, dollar values are more prone to measurement errors. Second, the frequency of the data (monthly) is known be exactly the frequency of decisions by the supermarket. Therefore, unlike most macroeconomic or industry-level datasets, the one here does not have the "time-disaggregation bias" identified by Kahn (1992). Third, industry- or even firm-level datasets have the problem that the items produced might not be of the same constitution over time. This is a well-known problem in GNP accounting, and is evidently a serious concern for economists (e.g., Blanchard (1983)'s automobiles over time). Fortunately, at the SKU level, items are assured to be qualitatively the same over the period studied. Fourth, retail also has the advantage that not only are the outputs qualitatively the same over time, the technology of production is not a confounding factor. There are no changing factor proportions and prices in the production of outputs. In the economics literature, for example, one has to appeal to the envelope theorem and smooth substitution possibilities to make for a convincing estimation (e.g., Kahn (1992)). Fifth, much of the contemporary theories on the bullwhip effect in the operations management literature pertains to the retail industry (e.g., Lee, et al. (1997b), Chen, et al. (2000)), so it seems appropriate to address this industry directly. Finally, I should mention that there is of course no unobserved time-invariant firm heterogeneity in using fixed effects with a panel dataset. But this is a double-edged sword, since the price paid is lower persuasive power in external validity. In summary, the dataset is a significant departure from those used in the literature.

## 5. Empirical Strategy

The empirical strategy is to couch the null hypotheses in terms of absence of the causes, seek suitable measures of the causes, and see if they explain amplification, thus rejecting the nulls. Conceptually, it seems more prudent and useful to be able to say that X is a cause, rather than X is not. Econometrically, it is difficult to comprehensively measure each cause, so the sizes of the tests would be too small if the nulls are couched as the presence of causes.

I begin by defining a "raw" amplification ratio for SKU $i$:

$$AMPLIFICATION\_RAW_i = Var(SALES_i) \,/\, Var(SUPPLIES_i)\ , \tag{1}$$

where $SALES_i$ is sales delivered to outlets and $SUPPLIES_i$ is supplies received from suppliers. Naturally, it is theoretically sounder to estimate (1) not just in reduced form parameters, but to recover structural parameters characterizing the technology. Given the focus of this paper on disentangling causes, I refer readers to Lai (2005) for a variance bounds test of a structural model. I now address the empirical challenges in identifying equation (1).

The raw definition will not be well-behaved if the $SALES_{it}$ and $SUPPLIES_{it}$ time series are trend stationary. Assuming that they are, I follow Blinder (1986) and use variances of the detrended series to construct two alternative definitions of amplificaiton. Specifically, I can regress for example, $SUPPLIES_{it}$, on a time trend and a set of 12 monthly dummies that account for additive seasonality, and construct the variance as:

$$(2) \qquad Var(SUPPLIES_i) = \sum_{m=1}^{12} \frac{(\hat{b}_{im} - \bar{b}_i)^2}{12} + MSE_i \ ,$$

where $\hat{b}_{im}$ is the predicted coefficient for the *m*th month dummy, $\bar{b}_i$ the mean of these the month coefficients, and $MSE_i$ the mean squared error of the regression. The first term on the right-hand-side can be interpreted as the seasonal deterministic component and the second the stochastic component, so that I obtain the two alternative definitions:

$$(3) \quad AMPLIFICATION\_SEASONAL_i = seasonal(SALES_i) \ / \ seasonal(SUPPLIES_i) \ ,$$

and

$$(4) AMPLIFICATION\_STOCHASTIC_i = stochastic(SALES_i) \ / \ stochastic(SUPPLIES_i) \ .$$

The standard errors can be calculated from a linear expansion of the amplification ratio in the parameters of the trend seasonal regressions. The distributions of these parameters are derived from Newey and West (1987) autocorrelation robust procedures. They are asymptotically normal, but only approximately so, since the exact distribution cannot be normal if the ratio is strictly positive.

If a series is not trend stationary, then the variance statistics are not well-defined, even with detrending. I then define the amplification ratio in other ways. One is to use differences:

$$(5) \qquad AMPLIFICATION\_RAW\_DIFF_i = Var(\Delta SALES_i) \ / \ Var(\Delta SUPPLIES_i) \ ,$$

where $\Delta$ is the first difference operator. This ratio is hard to interpret, but has been commonly used (e.g., Krane and Braun (1991)). Yet another definition is to exploit the cointegration that, in principle, should hold in a reduced-form relationship between $SUPPLIES_i$ and $SALES_i$, even if both are I(1) (integrated of order one). Therefore, I offer:

$$(6) \qquad AMPLIFICATION\_COINT_i = \zeta^2 \ ,$$

where $\zeta$ is the cointegration factor.

Recall the previous discussion that amplification might be defined as a ratio of maxima, rather than variances; hence this alternate:

(7) $AMPLIFICATION\_AMPLITUDE_i = max(SALES_i) / max(SUPPLIES_i)$ ,

By thinking of the raw, seasonal, stochastic, cointegration, and amplitude definitions as five basic ratios, I construct variations of these. First, for each, I can apply a difference operator, such as that applied to the raw ratio in equation (5). Second, I construct each variation at higher levels of aggregation: not SKU, but category, category-group, and firm. This allows me to check empirically what the impact of aggregation is. Third, for each, I construct parallel measures of amplification using dollar values rather than physical units, with the purpose of seeing how far off one could be from the other. Altogether, I construct 5 basic × 2 level/difference × 4 aggregations × 2 physical/dollar, or 80 variations. Given space limitations, I report analyses on only the basic ratios, but the summary statistics show all of them. In some analyses, I also report different aggregation levels to show the aggregation effect.

In constructing these definitions, I face some empirical issues. First, at the SKU level, one would hope that the accounting identity holds:

(8) $SUPPLIES_{it} = SALES_{it} + INVENTORY_{it} - INVENTORY_{i,t-1}$ ,

where $INVENTORY_{it}$ is the stock at the end of period $t$. Some slight variation might be expected due to lax controls or discipline in updating the database. To get a sense of the discrepancy, I run an OLS (ordinary least squares) estimation of (8), which should hold if the constant term is 0, the coefficients of the right-hand-side variables are 1, 1, and -1 respectively, the $R$-squared is 1, and any residual is white. The 3,745 SKU regressions pass almost all these tests. The $F$ statistics are all significant at the 0.00% level. The average $R$-squared is 99.56%. Portmanteau (Q) tests for white noise reject the null in only 50 SKUs, at the 1% level.

There is also smaller empirical issue of whether to use calendar or working days. Although the former is common, Fair (1989) argues that the latter is more accurate, but his context is in manufacturing production rather than retailing (and specifically, he looks at non-convexities in production technology). In any case, I check that the results here are invariant to both, but I present results using the more common calendar days.

Once the amplification ratios are estimated, the next step is to estimate the importance of each cause, using a model of the following basic form:

$$\log(AMPLIFICATION_i) = \mathbf{X}_i.\boldsymbol{\beta}_i + \varepsilon_i \,, \tag{9}$$

where $\boldsymbol{X}_i$ the vector of log covariates, including the causes and other necessary controls, and $\varepsilon_i$ is white noise. As before, although I write (9) at the SKU level, I will also estimate at the levels of category, category group, and firm, to see the effect of aggregation.

There are additional empirical issues in estimating (9). First, I have to be reasonably persuaded that the causes are exogenous. To allow for the possibility that *AMPLIFICATION* might respond to a contemporaneous $\boldsymbol{X}$, it will always be instrumented by variables in the lagged information set in my fixed effects estimations. In the discussion of how I measure each cause below, I will also describe the *a priori* economic reasons for why the measures are suitably exogenous. The second empirical issue is whether there is anything in the error term that might drive a spurious relationship between a cause and amplification. Again, while I will describe the specific situations for each cause below, I broadly control for these with seasonal dummies and trend terms. For higher level estimations, I also run fixed effects on category and category groups, and control for volumes. The third empirical issue is whether the SKUs might be correlated in some way. For example, greater amplification in one SKU due to trade promotions might be correlated (positively or negatively) with amplification in a related SKU. I deal with these issues by clustering at the respective levels – *e.g.*: at the category level for SKU regressions. The fourth issue is potential SKU-level heterogeneity. To get rid of at least the time-invariant components, I split the dataset into 6-month periods (with 5 months in the last), and use fixed effects in my estimations. This raises another issue of whether six months is a reasonable horizon. This is a standard topic in econometrics (e.g., Morck, et al. (1990) in choosing the horizon for measuring growth). Shorter periods provide more accurate representations of contemporaneous amplification but lower the power of the tests. Longer periods better capture variances, but are more susceptible to endogeneity problems. Having tried both 3- and 9-months without significant changes to the conclusions here, I present results using 6-month periods. The sixth issue is the potential effect of outliers. I re-do all the estimations with data winsorized at the 1% and 99%. The results are qualitatively unchanged, so I omit reporting those with winsorized data. Finally, there is a question of whether the decision cycle is really monthly, even though I am told that it is so. It seems reasonable to suppose that the company partially responds within the month to news about $SALES_{it}$. The main econometric implication of such intra-period adjustment is that in the cointegration regressions, $SUPPLIES_{it}$ might be correlated with the error term, which now might contain news about $SALES_{it}$. The solution is to instrument $SUPPLIES_{it}$ using variables in the period $t$-1 information set.

I now describe the causes individually: how I measure them and various empirical issues associated with each.

The first potential cause is from cost shocks. In my setting, these are most associated with trade promotions. Because there might be other shocks (e.g., Blanchard (1983) looks at labor strikes in the automobile production setting), I check Factiva for news articles using the search words "strike" or "unrest" (for labor costs), "transport*" or "logistic*" (transportation costs), "rent" and "commercial" that are associated with the Spanish retail industry in that period. Of the 46 articles found, the only ones remotely related to the setting is a 59-day strike in the sherry industry starting November 1991. I then look for sherry SKUs in my dataset by searching for any of the words in the SKU description: "Sherry", "Xerex", "Palomino", "Ximenez", "Muscat", "Moscatel", "Fino", "Flor", "Oloroso", "Manzanilla", "Amontillado", "Cortado", "Jerez." These are obtained from www.wikipedia.org and correspond to names like the town in which Spanish sherry is made (Jerez), the types of grapes (Palomino), fermentation styles (Fino). The dataset returns 52 SKUs, but on inspection, all by 13 are not really sherries but happen to have the words used in their descriptions. The 13 belong to two category groups

(Alcohol and Wine) and three categories (90116, 90118, 90501). However, it turns out that there is hardly any price variation on these sherry SKUs over the whole period of the dataset. Therefore, this can be safely ignored as a cost shock, although I use this as a natural experiment to investigate another cause (gaming in anticipation of shortage) below.

I then proceed to trade promotions as another kind of cost shocks, measuring their prevalence using two methods. The first is a count of the number of months in which official trade promotions occur. Since some price reductions from suppliers might not be officially so classified, the second method is to measure "abnormal" price reductions from the residuals of fixed effects regressions of supplier price on two lags, with month dummies and trends and no intercepts. The total price reduction over the 29 months is the sum of the monthly abnormal price reductions. Because it is likely that news about trade promotions come before orders are decided, I use these measures with one period ahead.

Empirically, I do not think there is an endogeneity issue of a trade promotion being initiated *because* of amplification. This seems implausible, and is different than the more plausible argument that a promotion might be started because of low volumes. Also more plausible is that promotions might be seasonal, which I control for with seasonal dummies.

The second possible cause of the bullwhip is the autocorrelation of demand in the presence of stock out costs. The weaker form of the theory also assumes the presence of lags. I rely on the derivation in Lee, et al. (1997b) (see also Chen, et al. (2000)), who assume that $SALES_{it}$ has AR(1) coefficient $\rho_i$ and the replenishment lag time is $L$. A Taylor expansion of their equation 3.5 is:

(10) $$AMPLIFICATION_i = 1 + 2\sigma_i^2/Var(SALES_i).[(1- \rho_i^{1+L})(1- \rho_i^{2+L})(\rho_i+O(\rho_i)^2] ,$$

where $\sigma_i^2$ is the variance of the residual in the $SALES_i$ series. In my setting, the 534-subsample has $L$ equal to 1, so this simplifies to:

(11) $$AMPLIFICATION_i = 1 + 2\sigma_i^2/Var(SALES_i).[(\rho_i+O(\rho_i)^2] .$$

Therefore, I use a specification with $log(\rho_i)$ as a covariate. This specification does not apply to the stronger form of the theory, in which a positive $L$ is not needed. Instead, I derive my specification from Kahn (1992), modifying it from his manufacturing setting to retail. For example, I set up ordering costs to be concave and I set factor costs to zero. His starting formulation, which is also that used by Lee, et al. (1997b) and Chen, et al. (2000), is that firms solve the following:

$$(12) \qquad Max\ E_{t-1}\left(\sum_{s=t}^{\infty} d_s^{s-t}\left[p_s SALES_s - c_s SUPPLIES_s\right]\right),$$

where $d_t$ is a discount factor, $p_t$ prices to customers, and $c_t$ costs from suppliers, and all are functions of time. In the appendix, I show that this leads to an econometric specification that is linear in the variance of markup, controlling for the covariance matrix of inventory, potential demand, and markup. The theory predicts significant positive coefficient for the variance of markup. Econometrically, it is easier to get the entire right-hand side of equation (16) together, which I will label simply as $X$. Therefore, the prediction is that $Var(X)$ is positively signed in a regression on amplification.

Fortunately, the dataset has direct measures of all three, even potential demand – *i.e.*, sales plus stock outs. Because this last is likely to be poorly measured, I instrument it with lagged values. Although stock out quantity is not the variable of interest here, its endogeneity can bias the estimation of $X$. Unlike the case with trade promotions, it seems more plausible there might be reverse causality here, through a learning mechanism. The possible story is that amplification is correlated with poor execution and stock outs, so that customers learn that with high amplification, they cannot depend on buying the volume they have bought previously, thus lowering demand autocorrelation. Since this is a learning mechanism, the lagged dependant variable specification would address the problem.

An important issue with the above estimation of the effect of $\rho_i$ is that the effect could have gone through a combination of rational and behavioral channels. Specifically, a behaviorist might argue that with under-reaction, managers systematically under-react to changes in $SALES_i$ because they overestimate $\rho_i$. It turns out this is irrelevant, because demand correlation is not a significant predictor of amplification, as the data will show.

The third possible cause is batching. The prediction is that in a cross section of SKUs, I should see that those which more batching experience higher amplification. Obviously, SKUs in the 534-subsample are not batched, so at a high level, I compare amplification between this subsample and the rest of the dataset. In addition, for robustness, I consider three types of batching: inter-temporal by SKU, across SKUs by supplier, and across retailers.

The idea of inter-temporal batching is that SKUs that are batched over time have longer duration between deliveries:

$$BATCHING\text{-}SKU_i = AVERAGE\text{-}MONTHS\text{-}BETWEEN\text{-}DELIVERIES_i\ ,$$

where the period concerned includes only those months in which there is either supplies received or sales delivered or both – *i.e.*, an active SKU. I also take care to censure the data at both ends of the period. To reduce measurement error, I instrument the measure with $WEIGHT_i$ / $MEAN\text{-}MARKUP_i$ , where $WEIGHT_i$ is the item weight of the SKU and $MEAN\text{-}MARKUP_i$ the total dollar markup over the period of the dataset. The heavier the item and lower the markup, the lower the likelihood that the SKU is batched over time.

The idea of cross-sectional batching across SKUs is that the retailer reviews with the same frequencies SKUs from the same supplier. For each SKU *i* supplied by supplier *S*, I construct:

$$BATCHING\text{-}SUPPLIER_{iS} = AVERAGE\text{-}MONTHS\text{-}BETWEEN\text{-}DELIVERIES_S\ ,$$

where the inter-delivery duration is now across all SKUs supplied by *S*. To get a measure of batching across retailers, I employ instrumental variables. I use the intuition that there are some SKUs that are so seasonal that many retailers tend to order them at about the same time. I choose two instruments: the *R*-squared of the seasonal components of supplies received and that for sales delivered. The idea is that, to the extent SKU quantities are predictable, all retailers would order at about the same time.

A third instrument is a dummy that indicates whether the SKU is a private label. It seems quite plausible that private labels, not carried by competing retailers, might be inversely correlated with *BATCHING-SUPPLIER*.

Again, there is a possibility that amplification influences batching, and again, I address this with lagged dependant variables. The harder empirical issues are with the last measure of batching. First, it could be that amplification and batching are both driven by calendar time, so any relationship between the first two might be spurious. The neater empirical approach is to use a natural experiment based on retailers changing the frequencies of their reviews, but such changes are hard to obtain. The mitigating issue here is that I use two other measures of batching, but the results are subject to this potential problem.

The fourth cause is gaming in the face of supply shortages. One opportunity is the natural experiment presented by the sherry strike mentioned earlier. I undertake an event analysis to test for gaming. Another approach is to use several instrumental variables. The first two are the (negative[2] of the) total returns from customers to the distribution center and returns from the center to suppliers. I do not have the reasons for these returns, and indeed, some of the returns might be for quality problems that come with production during shortages. Nevertheless, it seems plausible that on the whole, returns are (inversely) correlated, even if imperfectly, with shortages and much less with other causes. A third instrumental variable is (the reciprocal of) the number of SKUs actively sold in the category. The idea is that the more substitutes an SKU has, the less likely gaming occurs. A fourth instrument is the number of periods with stock outs. Finally, I construct an instrumental variable measuring "abnormal" growth in sales. The idea is that such abnormality might not be expected by suppliers, leading to next-period shortage. I measure per-month abnormal sales with the $\alpha_{it}$ in:

$$SALES_{it} = \alpha_{it} + \Sigma_{\phi}\beta_{\phi}SALES_{i,t-\phi} + \text{seasonal dummies} + \text{trend} + \eta_{it} \,,$$

---

[2] I take the negative rather than reciprocal of the return numbers because a number of them have zero values.

where $\phi$ is the number of lags in *SALES* to use and $\eta_{it}$ is assumed to be i.i.d. with zero mean. I set $\phi$ to 2 in the analyses reported here; unreported analyses with $\phi$ set to 1 or 3 produce qualitatively similar results. For SKU *i*, abnormal sales is defined as:

$$ABNORMAL\text{-}SALES_i = | \Sigma_t \, \eta_{it} \, | ,$$

where the summing is over the number of months in the period – *e..g,* in the cross-section, it is over all 29 months; in the panel, it is over the 6 months for each period (or 5, for the last). I undertake two-stage least-squares estimations with these instrumental variables. Although some of the instruments like the number of SKUs seem decisively exogenous, I undertake over-identification tests just to be sure that the instrumental variables approach is sound.

I now turn to measuring the behavioral causes. The first possible cause is perceptual bias. One type of bias is under-reaction to operational parameters like orders in the pipeline and lag time. Because my setting has only one-month or zero lags, I create a dummy $LAG_i$ for this. Recall that zero-lag SKUs by-pass the Sebastian distribution center because they are either perishable or have efficient suppliers. While it is possible that SKUs that show high amplification (suppliers becoming inefficient) might be pulled back into the center, this is not an issue in my dataset because there is no change in $LAG_i$ over time. Another perceptual bias is over-reaction. Here, I first ensure that there is over-reaction, by calculating the AR1 coefficient for $SUPPLIES\_RECEIVED_{it}$ (unlike the demand AR1, which is calculated for $SALES\_DELIVERED_{it}$). Then I regress amplification ratios on this AR1. In both the under- and over-reaction estimations, I control for sales *AR1*, which is likely to be correlated with this new supplies AR1.

The second behavioral cause is coordination risk. I use several instruments. The first is the diversity of technologies used to supply the retailer. The standard deviation of the item weights of the SKUs from each supplier is a proxy. Another instrument along the same line is the number of SKUs from each supplier. Again, the more fragmented, the higher the coordination risk. Both these measures need to control for the volume of business. The third instrument, therefore, is the importance of the retailer to the supplier. The intuition is that the less important, the more likely the supplier pays less attention to coordination issues. A measure of this is the (log of) the reciprocal of the total value from a supplier t the retailer. Finally, the fourth measure is "new" SKUs. I measure this with the average months since the last sale. To avoid truncation bias (those with short averages might still be new), I use only the data starting from month 7 in the dataset. I hasten to admit that these are rough proxies, but I add that the econometric technique requires only some correlation (and exogeneity which, if any, is managed with lagged dependent variables, and tested with over-identification tests).

For each cause, I first run regressions with just that cause, on amplification. The idea is to have the measures pick up as much as possible the true effects of the cause, *even* if this might be correlated with other causes. After the analyses by cause, I estimate the multivariate specification with all causes on the right-hand side. Now, I am controlling for correlations among the causes, expecting the pure partials to be lower. All the estimations are done with panel fixed effects where applicable, or cross-sections otherwise - *e.g.*, for $LAG_i$, in which a panel is meaningless since $LAG_i$ is time-invariant. I also run the regressions at the SKU, category, and category-group levels. Finally, I estimate with and without limitation to the 534-subsample. All these are on the various types of amplification ratios.

## 6. Results

Table 1 shows the summary statistics. The dataset is particularly useful for its level of detail. It also has rich amount of heterogeneity, in terms of value of goods, supply and demand levels, promotions, even weight.

In Table 2, I report the trend stationarity of each SKU's supplies and sales time series. The augmented Dickey-Fuller statistics suggest some minor unit root behavior. Cachon, et al. (2005) obtain a similar result, although at the higher industry level. The problem seems rather limited, especially after I incorporate trend terms and three lags, which are significant. As described earlier, I ensure that my amplification ratios are well-behaved by constructing various versions of the ratios, such as those using differences and cointegration.

In Table 3, panel (a), I document the various amplification ratios. For SKU level calculations, I run 3,745 × 4 regressions on the full dataset and 534 × 4 regressions on the subsample, where the 4 times comes from running the *SALES* and *SUPPLIES* series, each using physical units and dollar values. The results are estimated with fixed effects. Those done with OLS are clustered at the SKU, category-group, and category levels, and produce qualitatively unchanged results (unreported). The stochastic ratios tend to be larger than the seasonal ones, which is to be expected since seasonal sales and supplies might be more predictable. Some ratios seem wildly huge, but are limited to the SKU level and in the full dataset, where there are SKUs that are active for only a few months. The data also shows an intriguing empirical point about aggregation: there are *fewer* ratios above 1.0 at lower levels (like SKU), but the magnitude of the ratios are *bigger*. This suggests that there might be systematic biases in different ways, when looking at the bullwhip effect at high levels of aggregation. Another interesting point is that amplification is less evident using dollar values than physical units. Overall, the data shows evidence of the bullwhip effect. At the SKU level, at least 80% of physical ratios are over 1.0, for all definitions of amplification but cointegration.

In panel (b), I report the correlations between the different ratios. I show this for one example, but the result is similar when using dollar values (rather than physical units) on the 534-subsample (rather than the full dataset), at the category group level (rather than category or SKU). This example, however, seems optimistic. When I inspect the histogram of all the 240 coefficients at different levels, different datasets, etc., the mean is 0.556 and the median 0.550, although there is some comfort that all the correlations are non-negative. Therefore, it is worthwhile retaining the robustness checks of looking at a range of ratios.

In panel (c), I report the category groups with the lowest and highest raw amplification. Predictably, staples tend to have lower amplification, and goods for rarer purchases have higher ratios. Panel (d) shows the same, at the SKU level.

In the next few tables, I report the point estimates of the various causes. I show only selected results even though estimations are done on various combinations - *e.g.*: full sample vs. 534-subsample, physical vs. dollar units, SKU vs. category levels. Where the results are qualitatively similar, I do not present or discuss the variations.

In Table 4, I report estimates of cost shocks as a possible cause, using two measures of trade promotions for shocks. In panel (a), I report the results at the SKU level. The top half of the panel uses both the contemporaneous trade promotion as well as the period-ahead value. The columns are for the various amplification ratios as dependent variables. I omit the stochastic amplification ratio because it is under-identified in fixed effects panels: each period has only 6 observations corresponding to the 6 months in a half-year period, whereas the regressions to estimate it have many more variables such as the 12 monthly dummies. In any case, I am able to use the stochastic measure in cross-sectional regressions involving all months. The results are qualitatively the same and are not reported. In panel (a), only a few of the trade promotion variables are significantly positive. Some insignificant ones even have the wrong negative sign. The bottom half of the panel omits the contemporaneous trade promotion to get a sharper estimation. The *R*-squared does not drop appreciably, showing this to be a more reasonable specification. Formal Hausman tests confirm the same. The trade promotion variable is still only sparsely significant. Panels (b) and (c) are for category and category group levels. The point is to see if aggregation matters in attributing cause. The lack of significance is complete at these levels, suggesting that aggregation makes it even harder to detect cost shocks as a cause. Overall, the data provides no evidence that cost shocks are a significant driver. This is consistent with the more recent literature in macro-economics (e.g., Kahn (1992)).

In Table 5, I report the test of demand correlation as a possible cause. In panel (a), the SKU level estimations show that when the AR1 measure is used (recall that this assumes replenishment lags), the coefficient is negative, contrary to prediction. Two of them are statistically significant, although the economic significance of the point estimate is low. For example, a one standard deviation in log *AR1* is associated with only a 1.5% standard deviation change in log raw amplification. The coefficient on *Var*(*X*) (recall that this assumes no lags) fits that theory better, with mostly positive and statistically significant coefficients in the full sample. The economic significance is about an order of magnitude larger than that for *AR1* – for example, 13.5% change in log raw amplification. In panel (b), the results at the category level are that *Var*(*X*) behaves more in accord with theory. The economic significance is also slightly greater. For example, in the 534-subsample, a one standard deviation in log *Var*(*X*) is associated with an 87.5% change in log seasonal amplification. Overall, I interpret the result as generally very weak evidence that demand correlation is a causal factor. This is consistent with Cachon, et al. (2005), who find the same with industry level data.

In Table 6, I produce evidence for batching as a cause for the bullwhip effect. The coefficient of interest in panel (a), which reports SKU level estimations, is *BATCH*. Almost all the coefficients are correctly signed (positive) and statistically significant. The impact is also economically significant. For instance, in the fixed effects estimation, a one standard deviation in *BATCH_SKU* corresponds to an 18% standard deviation in log raw amplification and a one standard deviation in *BATCH_SUPPLIER* corresponds to a 21% in the same ratio. This significance is robust to however *AMPLIFICATION* or *BATCH* is defined, different estimation methodologies, and even different levels (such as category, which is shown in panel (b)).

In Table 7, I report the test for gaming. The first test is a natural experiment involving the sherry labor strike in the last two months of 1990. In panel (a), I first report the how the amplification ratios for sherry look like before, during, and after the strike. Time 1 corresponds to January through June 1990, time 2 to July through December 1990, and so on. If gaming were to occur, theory predicts that amplification might be the highest in time 2, assuming that all else being equal. Unfortunately, the data shows the opposite. The higher amplification ratios for sherry SKUs at times 1 and 3 might resuscitate the theory. For example, one could argue that either information about the strike is used by retailers way ahead (in time 1), or conversely, the effect of the strike (the shortage) is really much later than the strike itself, in which case amplification increases only after it (in time 3). This is especially appealing given that amplification seems to drop for all SKUs, and for all SKUs other than sherry. The rightmost part of panel (a) shows what if the impact of the strike is not just on sherry SKUs, but on all SKUs sharing the same category group or category as sherry SKUs. The pattern for this broader measure of the impact of the strike is similar to that for sherry SKUs themselves.

To get a more formal estimate of the impact of the strike, I report in panel (b) results from fixed effects estimations. The coefficient of interest is the interaction term of *SHERRY?* and *EVENT?*. Theory predicts that it should be positively signed, since retailers are to game in anticipation of shortage. The sign, however, is uniformly negative and significantly so. This is true for all amplification definitions, and whether I consider gaming only on sherry SKUs or all SKUs sharing the same category or category group as sherry SKUs. In case I might have mis-measured the timing of the shortage, I re-estimated the regression using time 1 and time 3 for *EVENT?*. The coefficient for the interaction term, shown in the bottom two rows, are statistically not different from zero except when I measure the impact broadly, on all SKUs sharing the same category group as sherry SKUs. But if I take this broad definition, then it seems hard to reconcile with the earlier evidence that time 2 coefficients are all negative, even just for sherry SKUs. Overall, I interpret this as insufficient evidence for gaming as a cause.

To make this point more robust, panel (c) shows that the coefficients on proxies for gaming are all insignificant; a few are even signed negatively. This is despite my using five instrumental variables, hoping to get to a tighter estimation. The use of this many instrumental variables is susceptible to at least one of them being exogenous, but the over-identifying test statistics mostly pass the tests. Just to show that this is not an artifact of being "too detailed," panel (d) reports the results at both the category and category group levels. Again, not much is statistically significant. All in all, I believe the data says that gaming is unlikely to be significant in this retail case.

In

Table 8, I show the results of the test for behavioral under-reaction as a cause. As usual, panel (a) reports SKU level estimations and panel (b), category level. The variable of interest is $LAG_i$, and it is positively signed and statistically significant, as predicted. I use two specifications, which control for *AR1* in different ways. In odd-numbered specifications limit observations to those SKUs in which the absolute value of *AR1* is smaller than 0.1 The even-numbered ones add log *AR1* as an explicit right-hand-side variable. The control for *AR1* is to account for the possible interaction between this behavioral cause due to lags and the rational, demand correlation cause due to the same. In any case, as I show earlier, demand correlation is not a major cause. More importantly, I account for interactions among all causes explicitly, later. The $LAG_i$ coefficient is economically significant. For example, one standard deviation increase in $LAG_i$ is associated with a 36% increase in the standard deviation of log raw amplification. Once again, the category level shows fewer cases with statistical significance, pointing to some aggregation effect. Overall, I interpret the results as some evidence that systematic under-reaction is a cause of the bullwhip.

In panel (c), I report the test of over-reaction as a possible cause. I first test whether supplies received are correlated with past supplies. In a SKU-by-SKU regression of $SUPPLIERS\_RECEIVED_{it}$ on its lagged value, I find that the AR1 coefficient has a mean of 0.14, a standard deviation of 0.31, and positively skewed at 5.51. Further, when I regress the sales AR1 on this supplies AR1, I get a positive 0.13 intercept, with a robust standard error of 0.005. In short, supplies are correlated, and are so even when sales are not. Returning to panel (c), I show that amplification is attenuated by the supplies AR1, whether using only those SKUs with absolute sales AR1 within 0.1 (odd-numbered models) or when I control for the log of the sales AR1 (even-numbered models). This is true for category-level evidence, too (unreported).

In Table 9, I report the results of testing coordination risk as a possible cause. In the first specification, $1/SUPPLIER\_VALUE_i$, $SUPPLIER\_SKUS_i$ and $SUPPLIER\_WEIGHT\_SD_i$ are all predicted to be positively signed. The result is generally consistent with these predictions, although the coefficients for $SUPPLIER\_WEIGHT\_SD_i$ are all insignificant. The economic significance is moderate. One standard deviation in $Log(SUPPLIER\_VALUE_i)$ is associated with 16% standard deviation in log raw amplification (this can be interpreted as an elasticity). In the second specification, *MTHS_SINCE_LAST_SALE*$_{it}$ is weakly consistent with the predicted positive sign; it is either significant and positive or insignificant. The category level estimation has the same qualitative result.

In Table 10, I report results of regressing amplification on all causes identified described above. Except for the lagged dependant variable and intercept, all variables are predicted to be positively signed. The conclusions thus far seem to hold up in the SKU level estimations in panel (a). Batching of all types and the behavioral variables all continue to be important. Demand correlation and gaming still are not. Cost shock is mixed. The category level estimations in panel (b) omit some of these variables because I run fixed effects, in which these variables are time-invariant and drop out. The result is decidedly more mixed at this level, although batching continues to generally stand up.

It is easier to see which causes are the most important using a simple technique, following Schmalensee (1985). In Figure 1, I run regressions on log raw amplification using different combinations of the cause. I lump the perceptual biases (under- and over-reaction) and coordination into "behavioral," for simpler presentation. Starting from the bottom, the chart shows that batching alone causes the bulk of the identified amplification. Adding others improve explanatory power, and it is easy to see that behavioral causes come next, followed by gaming and cost shocks. Demand correlation reduces the adjusted $R$-squared. To clarify, this analysis is based purely on explanatory power even though (as we have seen in Table 10), for example, cost shocks generally have better statistically significance and the predicted signs than gaming. Therefore, I interpret the figure as showing batching and behavioral causes to be important, but it has too low statistical power to say much about the other causes.

## 7. Alternative Interpretations and Conclusion

I have addressed questions of potential reverse causality, measurement errors, spurious relationships, and other empirical issues earlier. There are two remaining issues. The first is that of alternative interpretations: what if a variable could be interpreted as a different cause than the one it is purported to measure? As an example, suppose that returns to suppliers, now used as an instrument for gauging the severity of shortages (and therefore the opportunity for gaming) also picks up behavioral under-reaction. The story might be that past-period returns are automatically added by the supplier to the next delivery, and in the absence of back-logs from customers, high returns associated with high amplification might be due to the retailer's ignoring returns. This is certainly plausible, but my empirical strategy has been to find more than one way to capture a particular cause. Therefore, it cannot be the case that all the five measures of gaming reported in Table 7 are susceptible to this interpretation.

The second, more difficult, issue is that I could have been so stringent in the measures that the tests have overly low power. In the regression with all the causes, the *R*-squared is a respectable but low 14%. The low *R*-squared makes it difficult to conclude that the other causes are not present, just that they cannot be detected here. All the data show is that batching and behavioral causes are present, not that they are the most important.

Going forward, higher power tests would be a clear research priority. It would be intriguing if, despite better tests, the *R*-squared continues to be low. That would cry for new explanations for the bullwhip effect.

Another interesting research direction is to link the causes to financial outcomes. Chen and Samroengraja (2004) show a model in which fixing the bullwhip effect may not lead to reduction in total costs of the supply chain. It would be profitable to test that empirically.

Finally, a natural progression would be to see if the above results hold up in different settings, whether in other retailers or in say, manufacturing.

## References


Aguirregabiria, V. "The Dynamics of Markups and Inventories in Retailing Firms," *Review of Economic Studies*, 66(1999), 275-308.

---. *Strategic Stockouts in Supermarkets*: 27. Boston: Boston University Economics Department, 2004

Anderson, E., C. Fine, and G. Parker. "Upstream Volatility in the Supply Chain: The Machine Tool Industry as a Case Study," *Production and Operations Management*, 9(2000), 239-261.

Baganha, M., and M. A. Cohen. "The Stabilizing Effect of Distribution Center Inventory," *Operations Research*, 46(1998), S72-S83.

Bivin, D. G. "Bunching in the Production Process," *Economics Letters*, 50(1996), 259-263.

Blanchard, O. "The Production and Inventory Behavior of the American Automobile Industry," *Journal of Political Economy*, 91(1983), 365-400.

Blinder, A. S. "Can the Production Smoothing Model of Inventory Behavior Be Saved?," *Quarterly Journal of Economics*, 101(1986), 431-454.

Blinder, A. S., and L. J. Maccini. "Taking Stock: A Critical Assessment of Recent Research on Inventories," *Journal of Economic Perspectives*, 5(1991), 73-96.

Cachon, G. P. "Managing Supply Chain Demand Variability with Scheduled Ordering Policies," *Management Science*, 45(1999), 843-856.

Cachon, G. P., and M. A. Lariviere. "Capacity Choice and Allocation: Strategic Behavior and Supply Chain Performance," *Management Science*, 45(1999a), 1091-1108.

---. "Capacity Allocation Using Past Sales: When to Turn-and-Earn," *Management Science*, 45(1999b), 685-703.

Cachon, G. P., G. Schmidt, and T. Randall. *Empiricism in Supply Chain Management Research*. Paper presented at the Supply Chain Roundtable, Philadelphia, PA, 2004.

Cachon, G. P., T. Randall, and G. Schmidt. *In Search of the Bullwhip Effect*. Philadelphia, PA: Wharton School, University of Pennsylvania, 2005

Caplin, A. "The Variability of Aggregate Demand with (S, S) Inventory Policies," *Econometrica*, 53(1985), 1395-1410.

Carlos, F. D. "On the Stability of Supply Chains," *Operations Research*, 52(2004), 909.

Cetinkaya, S., and C.-Y. Lee. "Stock Replenishment and Shipment Scheduling for Vendor-Managed Inventory Systems," *Management Science*, 46(2000), 217-232.

Chatfield, D. C., J. G. Kim, T. P. Harrison, and J. C. Hayya. "The Bullwhip Effect-Impact of Stochastic Lead Time, Information Quality, and Information Sharing: A Simulation Study," *Production and Operations Management*, 13(2004), 340.

Chen, F., Z. Drezner, J. K. Ryan, and D. Simchi-Levi. "Quantifying the Bullwhip Effect in a Simple Supply Chain: The Impact of Forecasting, Lead Times, and Information," *Management Science*, 46(2000), 436.

Chen, F., and R. Samroengraja. "Order Volatility and Supply Chain Costs," *Operations Research*, 52(2004), 707-722.

Clark, J. M. "Business Acceleration and the Law of Demand: A Technical Factor in Economic Cycles," *Journal of Political Economy*, 25(1917), 217-235.

Croson, R., and K. Donohue. "Impact of Pos Data Sharing on Supply Chain Management: An Experimental Study," *Management Science*, 12(2003), 1-11.

Croson, R., K. Donohue, E. Katok, and J. D. Sterman. *Order Stability in Supply Chains: Coordination Risk and the Role of Coordination Stock*, *MIT Sloan Working Paper 4513-04*: 32. Cambridge, MA: MIT Sloan School of Management, 2004

Croson, R., and K. Donohue. "Behavioral Causes of the Bullwhip Effect and the Observed Value of Inventory Information," *Management Science*(forthcoming), 37.

Daganzo, C. F. "On the Stability of Supply Chains," *Operations Research*, 52(2004), 909-921.

Dejonckheere, J., S. M. Disney, M. R. Lambrecht, and D. R. Towill. "Measuring and Avioiding the Bullwhip Effect: A Control Theoretic Approach," *European Journal of Operational Research*, 147(2003), 567.

Fair, R. C. "The Production-Smoothing Model Is Alive and Well," *Journal of Monetary Economics*, 24(1989), 353-370.

Fisher, M. L., and A. Raman. "Reducing the Cost of Demand Uncertainty through Accurate Response to Early Sales," *Operations Research*, 44(1996), 87-99.

Forrester, J. *Industrial Dynamics: A Major Breakthrough for Decision Makers*, *Harvard Business*

*Review*, Vol. 36: 37-66, 1958

Ge, Y., J. B. Yang, N. Proudlove, and M. Spring. "System Dynamics Modelling for Supply-Chain Management: A Case Study on a Supermarket Chain in the Uk," *International Transactions in Operational Research*, 11(2004), 495.

Gilbert, K. "An Arima Supply Chain Model," *Management Science*, 51(2005), 305–310.

Gorman, M. F., and J. I. Brannon. "Seasonality and the Production-Smoothing Model," *International Journal of Production Economics*, 65(2000), 173-178.

Graves, S. C. "A Single-Item Inventory Model for a Nonstationary Demand Process," *Manufacturing & Service Operations Management*, 1(1999), 50-61.

Hammond, J. *Barilla Spa, HBS Case 694046*: 23. Boston: Harvard Business School, 1994

Holt, C. C., F. Modigliani, J. Muth, and H. A. Simon. *Planning Production, Inventories and Work Force*. Englewood Cliffs, NJ, Prentice-Hall, 1960

Holt, C. C., F. Modigliani, and J. P. Shelton. "The Transmission of Demand Fluctuations through a Distribution and Production System, the Tv-Set Industry," *Canadian Journal of Economics*, 1(1968), 718-739.

Humphreys, B. R. "The Behavior of Manufacturers' Inventories: Evidence from Us Industry Level Data," *International Journal of Production Economics*, 71(2001), 9-20.

Kahn, J. A. "Inventories and the Volatility of Production," *American Economic Review*, 77(1987), 667-679.

---. "Why Is Production More Volatile Than Sales? Theory and Evidence on the Stockout-Avoidance Motive for Inventory-Holding," *Quarterly Journal of Economics*, 107(1992), 481-510.

Kleijnen, J. P. C., and M. T. Smits. "Performance Metrics in Supply Chain Management," *The Journal of the Operational Research Society*, 54(2003), 507.

Krane, S. D., and S. N. Braun. "Production Smoothing Evidence from Physical-Product Data," *Journal of Political Economy*, 99(1991), 558-581.

Lai, R. *What's Inventory For? A Variance-Bounds Test of a Multi-Purpose Inventory Model*. Boston: Harvard Business School, 2005

Lee, H. L., V. Padmanabhan, and S. Whang. "The Bullwhip Effect in Supply Chains," *Sloan Management Review*, 38(1997a), 93.

---. "Information Distortion in a Supply Chain: The Bullwhip Effect," *Management Science*, 43(1997b), 546.

---. "Comments on "Information Distortion in a Supply Chain: The Bullwhip Effect"," *Management Science*, 50(2004), 1875.

Mack, R. P. "The Process of Capital Formation in Inventories and the Vertical Propagation of Business Cycles," *Review of Economics and Statistics*, 35(1953), 181-198.

Metzler, L. A. "The Nature and Stability of Inventory Cycles," *Review of Economic Statistics*, 23(1941), 113-129.

Miron, J. A., and S. P. Zeldes. "Seasonality, Cost Shocks, and the Production Smoothing Model of Inventories," *Econometrica*, 56(1988), 877-908.

Morck, R., A. Shleifer, and R. Vishny. "The Stock Market and Investment: Is the Market a Sideshow?," *Brookings Papers on Economic Activity*, 2(1990), 157-215.

Mosser, P. A. "Trade Inventories and (S,S)," *Quarterly Journal of Economics*, 106(1991), 1267-1286.

Newey, W. K., and K. D. West. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(1987), 703-708.

Schmalensee, R. "Do Markets Differ Much?," *American Economic Review*, 75(1985), 341-351.

Steckel, J. H., S. Gupta, and A. Banerji. "Supply Chain Decision Making: Will Shorter Cycle Times and Shared Point-of-Sale Information Necessarily Help?," *Management Science*, 50(2004), 458-464.

Sterman, J. D. "Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment," *Management Science*, 35(1989), 321-339.

Summers, L. H. "Retail Inventory Behavior and Business Fluctuations: Discussion," *Brookings Papers on Economic Activity*, 2(1981), 513-517.

Terwiesch, C., Z. J. Ren, T. H. Ho, and M. A. Cohen. "An Empirical Analysis of Forecast Sharing in the Semiconductor Equipment Supply Chain," *Management Science*, 51(2005), 208-220.

Tversky, A., and D. Kahneman. "Availability: A Heuristic for Judging Frequency and Probability," *Cognitive Psychology*, 5(1973), 207-232.

West, K. D. "The Sources of Fluctuations in Aggregate Inventories and Gnp," *Quarterly Journal of Economics*, 105(1990), 939-971.

# Appendix

Derivation 1 – Specification for Testing Stock out Costs as a Cause

Start with the per-period objective as in (12). Since stock outs are possible, write:

(13) $$SALES_t = min\ [\ INVENTORY_{t-1} + SUPPLIES_t\ ,\ DEMAND_t]\ ,$$

where $INVENTORY_{t-1}$ is at the end of period $t$ -1 and $DEMAND_t$ is potential demand conditional on price, and is assumed to be stochastic and exogenous. It has two components, $SEASONAL_t$ and $STOCHASTIC_t$ :

(14) $$DEMAND_t = SEASONAL_t + E_{t-1}(STOCHASTIC_t) + \xi_t\ ,$$

and $\xi_t$ is white with a monotonic distribution $\Phi$ and mean 1. Therefore, the Euler equation of (12) is:

(15) $$(p_t - c_t)(1 - \Phi[(INVENTORY_{t-1} + SUPPLIES_t)/E_t(DEMAND_t)] =$$

$$(c_t\ - d_t.E_tc_{t+1})\Phi[(INVENTORY_{t-1} + SUPPLIES_t)/E_t(DEMAND_t)]\ .$$

Using this to solve (12), I get:

(16) $$SUPPLIES_t{}^* = [INVENTORY_{t-1} + E_t(DEMAND_t)].\Phi^{-1}[\ \frac{p_t - c_t}{p_t - d_t E_t c_{t+1}}\ ]\ ,$$

where the right multiple on the right hand side can be interpreted as the stock out cost, measured using the markup. Controlling for variances and covariances of inventory and potential demand, the amplification ratio is then a function of the variance of this stock out cost.

# Figure 1 – Contribution of Explanatory Power

The percentages are adjusted $R$-squared's from OLS estimations using the full sample in a cross-section. The dependant variable is log raw amplification. The measures for the various explanatory variables are: $\mathit{PROMO\_COUNT}_t$ for cost shock, $Log(\mathit{AR1}_t)$ for demand correlation, $\mathit{BATCH\_SKU}_i$, $\mathit{BATCH\_SUPPLIER}_i$, and $\mathit{BATCH\_RETAILER}_i$ for batching, reciprocal of $\mathit{NUM\_OF\_BRANDS\_IN\_CAT}_i$ for gaming, $\mathit{LAG}_i$, Log(1/$\mathit{SUPPLIER\_VALUE}_i$), and $\mathit{SUPPLIER\_SKUS}_i$ for behavorial. The estimation is with robust standard errors and clustered at the SKU level. All estimations have $p$ values that are practically zeroes. The dark boxes show the highest $R$-squared regression with that number of variables in the specification.

| |
|---|
| Cost shock<br>Demand<br>Batch<br>Gaming<br>Behav |
| 13.7% |

| | | | | |
|---|---|---|---|---|
| Demand<br>Batch<br>Gaming<br>Behav | Cost shock<br>Batch<br>Gaming<br>Behav | Cost shock<br>Demand<br>Gaming<br>Behav | Cost shock<br>Demand<br>Batch<br>Behav | Cost shock<br>Demand<br>Batch<br>Gaming |
| 13.7% | 13.8% | 2.3% | 13.6% | 12.9% |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Batch<br>Gaming<br>Behav | Demand<br>Gaming<br>Behav | Demand<br>Batch<br>Behav | Demand<br>Batch<br>Gaming | Cost shock<br>Gaming<br>Behav | Cost shock<br>Batch<br>Behav | Cost shock<br>Batch<br>Gaming | Cost shock<br>Demand<br>Behav | Cost shock<br>Demand<br>Gaming | Cost shock<br>Demand<br>Batch |
| 13.8% | 1.9% | 13.7% | 13.0% | 2.7% | 13.7% | 13.1% | 2.3% | 0.9% | 12.9% |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Cost shock<br>Demand | Cost shock<br>Batch | Cost shock<br>Gaming | Cost shock<br>Behav | Demand<br>Batch | Demand<br>Gaming | Demand<br>Behav | Batch<br>Gaming | Batch<br>Behav | Gaming<br>Behav |
| 0.9% | 13.0% | 1.0% | 2.8% | 12.9% | 0.0% | 1.9% | 13.1% | 13.7% | 2.4% |

| | | | | |
|---|---|---|---|---|
| Cost shock | Demand | Batch | Gaming | Behav |
| 1.1% | 0.0% | 13.0% | 0.0% | 2.5% |

## Table 1 – Summary Statistics

The data is for from Sebastian, a Spanish supermarket. This data is for SKUs passing through the supermarket's distribution center, from suppliers to outlets. The frequency is monthly, and the dataset covers the 29 months from January 1990 through May 1992. 1,000 peseta is roughly US$7.40 as of August 2005 (after conversion through the Euro, which has replaced the peseta).

| | N | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| SKUs | 3745 | | | | |
| Categories | 408 | | | | |
| Category groups | 76 | | | | |
| Number of suppliers | 494 | | | | |
| Sales delivered (units/month) | 108,605 | 1,709.64 | 9,907.52 | 0 | 572,316 |
| Sales demanded (units/month) | 108,605 | 1,857.78 | 10,590.81 | 0 | 592,728 |
| Sales, list price (pta/unit) | 58,174 | 250.96 | 366.56 | 5.00 | 20,917.00 |
| Sales, promo price (pta/unit) | 20,984 | 206.70 | 270.25 | 0.01 | 19,324.23 |
| Sales returned by customers (units/month) | 108,605 | 22.71 | 287.54 | 0 | 41,548 |
| Stock out dummy | 15,486 | 0.29 | 0.46 | 0 | 1 |
| Supplies received | 108,605 | 1,735.82 | 10,641.59 | 0 | 584,052 |
| Supplies returned to supplier | 108,605 | 6.41 | 293.49 | 0 | 49,840 |
| Cost of supplies, list | 76,996 | 221.96 | 393.24 | 3.43 | 16,524.00 |
| Trade promotion dummy | 108,605 | 0.19 | 0.39 | 0 | 1 |
| Month-end inventory (units) | 108,605 | 1,016.22 | 3,810.10 | -23,211 | 179,282 |
| Private label dummy | 108,605 | 0.04 | 0.19 | 0 | 1 |
| Active months (out of 29) | 108,605 | 15.53 | 8.81 | 1 | 29 |
| Number of brands active in category | 108,605 | 10.06 | 8.52 | 0 | 46 |
| Weight per item | 92,771 | 2.69 | 13.74 | 0.00 | 374.00 |

## Table 2 – Trend Stationarity

The augmented Dickey-Fuller regression is run 4 x 3,745 times, with two specifications for each SKU's supplies or sales time series (hence the four columns below). The *p*-values are the MacKinnon approximates.

| | *SUPPLIES* | | *SALES* | |
|---|---|---|---|---|
| | No trend, 1 lag | Trend, 3 lags | No trend, 1 lag | Trend, 3 lags |
| Average *p*-value | 0.09 | 0.35 | 0.17 | 0.41 |
| Percent of SKUs with unit root at 5% significance | 25% | 17% | 48% | 12% |

## Table 3 – Amplification Ratios

Panel (a) – Summary

The general formula amplification ratio is *Var*($SALES_i$) / *Var*($SUPPLIES_i$) for levels (left half of page) and *Var*($\Delta SALES_i$) / *Var*($\Delta SUPPLIES_i$) for differences. Consider just $SALES_i$ to illustrate the construction of the different ratios; the construction of variances for $SUPPLIES_i$, $\Delta SALES_i$ and $\Delta SUPPLIES_i$ are analogous for the other ratios, except the cointegration one (see below). For the raw ratio, take the variances of $SALES_{it}$, directly from its time series. For the seasonal ratio, the variance is that of the seasonal dummies $b_i$ from a regression of $SALES_{it}$ on a time trend and monthly dummies. For the stochastic ratio, the variance is the mean squared error of the same regression. For the amplitude ratio, it is the maximum of the time series. The cointegration ratio is square of the cointegration parameter in a regression of $SALES_{it}$ on $SUPPLIES_{it}$, with a time trend. The ratios are constructed at four levels: pooling all observations together, or by category group, category, or SKU. Each of these are further done with physical units or dollar units, where the latter can also be thought of as physical units weighted by actual price (actual price is either the promotional or list price, whichever is lower). A further distinction is made: the ratios are calculated either using the entire dataset of 3745 SKUs, or the subsample containing the 534 SKUs which are actively sold in every of the 29 months in the dataset. "Mean" is the arithmetic average of the ratio (e.g., of all ratios calculated category). "% > 1.0" is the percentage (say of all ratios calculated by category) that exceeds 1.0. All regressions are done with fixed effects, corrected with heteroscedastic robust standard errors and are clustered around category groups, categories, or SKUs, respectively.

| | | | | **Levels** | | | | | **Differences** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Raw | Seasonal | Stochastic | Coint | Amplitude | Raw | Seasonal | Stochastic | Coint | Amplitude |
| **Pooled** | Phys | All | Mean | 1.15 | 1.46 | 1.15 | 1.04 | 1.02 | 1.63 | 1.56 | 1.63 | 0.98 | 1.02 |
| | units | 534-subsample | Mean | 1.23 | 1.42 | 1.23 | 1.07 | 1.19 | 1.79 | 1.45 | 1.79 | 1.09 | 1.19 |
| | Dollar | All | Mean | 0.96 | 1.07 | 0.96 | 1.06 | 1.07 | 1.63 | 1.56 | 1.63 | 0.98 | 1.07 |
| | units | 534-subsample | Mean | 1.00 | 0.99 | 1.00 | 1.11 | 1.07 | 1.79 | 1.45 | 1.79 | 1.09 | 1.07 |
| **Cat- groups** | Phys | All | Mean | 1.65 | 1.54 | 1.64 | 1.03 | 1.39 | 2.76 | 1.66 | 2.76 | 1.03 | 1.39 |
| | units | | % > 1.0 | 100.0% | 95.0% | 100.0% | 62.0% | 88.0% | 100.0% | 95.0% | 100.0% | 62.0% | 88.0% |
| | | 534-subsample | Mean | 2.3 | 1.65 | 2.63 | 1.36 | 1.49 | 4.46 | 1.95 | 4.79 | 1.36 | 1.49 |
| | | | % > 1.0 | 100.0% | 100.0% | 98.0% | 58.0% | 88.0% | 100.0% | 93.0% | 98.0% | 58.0% | 88.0% |
| | Dollar | All | Mean | 1.35 | 1.16 | 1.35 | 0.96 | 1.32 | 2.76 | 1.66 | 2.76 | 0.96 | 1.32 |
| | units | | % > 1.0 | 83.0% | 67.0% | 83.0% | 43.0% | 86.0% | 100.0% | 95.0% | 100.0% | 43.0% | 86.0% |
| | | 534-subsample | Mean | 2.43 | 1.5 | 3.09 | 1.38 | 1.34 | 4.46 | 1.95 | 4.79 | 1.38 | 1.34 |
| | | | % > 1.0 | 90.0% | 65.0% | 89.0% | 40.0% | 75.0% | 100.0% | 93.0% | 98.0% | 40.0% | 75.0% |
| **Categories** | Phys | All | Mean | 2.17 | 1.76 | 2.26 | 1.16 | 1.53 | 4.12 | 1.98 | 4.16 | 1.16 | 1.53 |
| | units | | % > 1.0 | 99.0% | 97.0% | 99.0% | 63.0% | 89.0% | 99.0% | 95.0% | 99.0% | 63.0% | 89.0% |
| | | 534-subsample | Mean | 2.3 | 1.69 | 2.58 | 1.25 | 1.49 | 4.29 | 1.96 | 4.61 | 1.25 | 1.49 |
| | | | % > 1.0 | 99.0% | 95.0% | 97.0% | 60.0% | 89.0% | 99.0% | 96.0% | 97.0% | 60.0% | 89.0% |
| | Dollar | All | Mean | 3.4 | 1.45 | 2.09 | 1.06 | 1.39 | 4.12 | 1.98 | 4.16 | 1.06 | 1.39 |
| | units | | % > 1.0 | 84.0% | 72.0% | 82.0% | 44.0% | 84.0% | 99.0% | 95.0% | 99.0% | 44.0% | 84.0% |
| | | 534-subsample | Mean | 2.41 | 1.55 | 2.84 | 1.21 | 1.34 | 4.29 | 1.96 | 4.61 | 1.21 | 1.34 |
| | | | % > 1.0 | 86.0% | 67.0% | 86.0% | 44.0% | 80.0% | 99.0% | 96.0% | 97.0% | 44.0% | 80.0% |

Continued…

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **SKUs** | Phys units | All | Mean | 6.06 | 1.86 | 6.7 | 4.19 | 1.76 | 8.86 | 2.17 | 9.28 | 4.19 | 1.76 |
| | | | % > 1.0 | 95.0% | 94.0% | 94.0% | 58.0% | 87.0% | 93.0% | 92.0% | 92.0% | 58.0% | 87.0% |
| | | 534-subsample | Mean | 6.53 | 1.9 | 6.9 | 3.52 | 1.82 | 9.55 | 2.24 | 10.45 | 3.52 | 1.82 |
| | | | % > 1.0 | 93.0% | 92.0% | 93.0% | 55.0% | 85.0% | 93.0% | 92.0% | 92.0% | 55.0% | 85.0% |
| | Dollar units | All | Mean | 5.95 | 2.35 | 346.67 | 213.67 | 1.56 | 8.86 | 2.17 | 9.28 | 213.67 | 1.56 |
| | | | % > 1.0 | 84.0% | 76.0% | 83.0% | 50.0% | 79.0% | 93.0% | 92.0% | 92.0% | 50.0% | 79.0% |
| | | 534-subsample | Mean | 6.08 | 1.87 | 43.86 | 26.44 | 1.62 | 9.55 | 2.24 | 10.45 | 26.44 | 1.62 |
| | | | % > 1.0 | 83.0% | 73.0% | 83.0% | 50.0% | 79.0% | 93.0% | 92.0% | 92.0% | 50.0% | 79.0% |

Panel (b) Correlations Among Amplification Ratios

This is an example for the category-group ratios, for the entire dataset, using physical units.

| | Raw | Seasonsal | Stochastic | Cointegration | Amplitude |
|---|---|---|---|---|---|
| Raw | 1.00 | | | | |
| Seasonal | 0.50 | 1.00 | | | |
| Stochastic | 0.99 | 0.48 | 1.00 | | |
| Cointegration | 0.43 | 0.21 | 0.45 | 1.00 | |
| Amplitude | 0.80 | 0.39 | 0.80 | 0.35 | 1.00 |

Panel (c) Example Category Groups at Extreme Amplification Ratios

The ratios are calculated for physical units, using levels, and on the entire dataset.

| Raw ratio | | Seasonal ratio | | Stochastic ratio | |
|---|---|---|---|---|---|
| 1.05 | Hydrogen peroxide | 0.86 | Custards, gelatins | 1.04 | Eggs |
| 1.06 | Eggs | 0.93 | Soap | 1.06 | Hydrogen peroxide |
| 1.09 | Milk | 0.97 | Fruits | 1.09 | Milk |
| 1.10 | Tissue | 0.99 | Champagne | 1.11 | Tissue |
| 1.12 | Sugar | 1.07 | Confectionary | 1.12 | Sugar |
| … | … | … | … | … | … |
| 3.07 | Condoms | 2.38 | Spot remover | 2.78 | Condoms |
| 3.35 | Paint, brushes | 2.38 | Condoms | 3.22 | Paint, brushes |
| 3.55 | Paper | 2.50 | Pet food | 3.48 | Paper |
| 4.40 | Stationery | 2.55 | Bicarbonates | 4.36 | Stationery |
| 4.47 | Brushes, combs, mirrors | 2.78 | Cologne, after-shave | 4.49 | Brushes, combs, mirrors |

(d) Example SKUs with Extreme and Median Amplification Ratios

The ratios are calculated for physical units, using levels, and on the entire dataset. Translation done at www.freetranslation.com.

| Raw ratio | | Seasonal ratio | | Stochastic ratio | |
|---|---|---|---|---|---|
| 0 | 1G OLIVE OIL COOSUR CARTON | 0 | 1G OLIVE OIL COOSUR CARTON | 0 | 1G OLIVE OIL COOSUR CARTON |
| 0 | ALMENDA TOST. THE AGUALI 100G | 0 | ALMENDA TOST. THE AGUALI 100G | 0 | ALMENDA TOST. THE AGUALI 100G |
| 0 | ALMOND TOSTADA LIGON 150 G | 0 | ALMOND TOSTADA LIGON 150 G | 0 | ALMOND TOSTADA LIGON 150 G |
| 0 | ATUN CLEAR GRANDS HOTELS 320 | 0 | ATUN CLEAR GRANDS HOTELS 320 | 0 | ATUN CLEAR GRANDS HOTELS 320 |
| 0 | CHAMPU + LOCION GREEN CROSS | 0 | CHAMPU + LOCION GREEN CROSS | 0 | CHAMPU + LOCION GREEN CROSS |
| 0 | CHAMPU you HANG YOU LIMON 400 CC | 0 | CHAMPU you HANG YOU LIMON 400 CC | 0 | CHAMPU you HANG YOU LIMON 400 CC |
| 0 | CHICKPEA HOSTAL 500 GRAMS | 0 | CHICKPEA HOSTAL 500 GRAMS | 0 | CHICKPEA HOSTAL 500 GRAMS |
| 0 | CHICORY MULLERE 200 GRS. | 0 | CHICORY MULLERE 200 GRS. | 0 | CHICORY MULLERE 200 GRS. |
| 0 | CHIPIRON BACKFILL FLORITA 115 | 0 | CHIPIRON BACKFILL FLORITA 115 | 0 | CHIPIRON BACKFILL FLORITA 115 |
| 0 | COLOGNE FA LIT. + DESOD FA 320 | 0 | COLOGNE FA LIT. + DESOD FA 320 | 0 | COLOGNE FA LIT. + DESOD FA 320 |
| … | … | … | … | … | … |
| 2.06 | MUSTARD COUSIN 285 GRS. | 1.56 | FONTANEDA MARIA PTES 200 GRS | 2.29 | ROLE ALUMINUM S. F. 16 MTS |
| 2.06 | LIQUOR MELON M. BRIZARD 3/4 | 1.56 | ESPARR-LETICIA YOLKS STANDAR | 2.29 | ESPARR. LETICIA 12/16 FM 1/2 |
| 2.06 | PEAS LETICIA 390 GRS. | 1.56 | SAUCE COCKTAIL YBARRA 215 GR | 2.29 | NORIT I LIQUIDATE GREEN MAQ.LTRO |
| 2.06 | PA¥OR COOKS WEEKLY | 1.56 | I WOUND BABY 250G RICE HAKE | 2.29 | PROTECTIVE FEMINA 30 you UNITE |
| 2.06 | ESPARRAGOS CIDACOS 9/12 KG. | 1.56 | EXPRESION REFLECTED GOLDEN | 2.30 | REGLERO PASTES OF YOU 525 GRS |
| 2.06 | RAICES-TIPS HAIR CREAM 450 | 1.56 | WINE C. V. N. AND 3 TO¥OR | 2.30 | ATUN CLEAR MIAU RO-70 |
| 2.06 | EVAX NUOVA EXTRAPL. 20 6322 | 1.56 | CREAM NOCILLA 2GUST 200 TARR | 2.30 | MINT POMPADOUR 10 ENVELOPES |
| 2.06 | CLEANING SANIGEL 850 | 1.56 | SPRING ONIONS RIOVERDE 200 | 2.30 | JABON I LIQUIDATE NENUCO LITER |
| 2.06 | LIMPIAHOGAR VIM SUPERFRESH L | 1.56 | MACED. VEGETABLES MARTINEZ 390 | 2.30 | CHAMPU NEUTER BALANCE 500 CC |
| 2.06 | CHOIR BOER MOKA 300 GRS. | 1.56 | PALILLOS FLAT CASE | 2.30 | MOP VILEDA WITH SPONGE |
| … | … | … | … | … | … |
| 78.91 | ATUN CLEAR GROVEMAR ACT.71,5 | 8.52 | COLOGNE PATH 110+ATOMIZ.60 | 93.01 | DELIAL SOLAR F8 BALSAMO 200G |
| 91.25 | GRACE T.YOLK TOSTADA 300G | 8.53 | JUICE COFRUTOS ORANGE BOT160 | 94.59 | ATUN CLEAR GROVEMAR ACT.71,5 |
| 92.77 | DELIAL SOLAR F8 BALSAMO 200G | 8.80 | S.F SEEDS OIL. LITER | 96.72 | SQUID TR.GROVEMAR ACT.2X65 |
| 133.81 | CHAMPAGNE WIDOWED CLIQUOT 3/4 | 8.86 | ALMEND. LEAVE/FRI BORGES 150 B | 117.09 | JUICE COFRUTOS ORANGE BOT160 |
| 144.00 | LIQUOR MELOCOTON SPREAD LIT. | 9.11 | ALMOND BORGES TOST. 150 B. | 144.78 | LIQUOR MELOCOTON SPREAD LIT. |
| 167.05 | ESP.POTE COLORANT you EXCEED | 9.26 | COLOGNE DARLING 100 ML. | 291.74 | ESP.POTE COLORANT you EXCEED |
| 400.00 | ALCACHOF.CHOSEN 16/20 1/2 | 9.58 | SQUID TR.GROVEMAR ACT.2X65 | 400.00 | ALCACHOF.CHOSEN 16/20 1/2 |
| 1296.30 | ALMIDON MIDOVINIL | 10.33 | ATUN CLEAR GROVEMAR ACT.71,5 | 1111.11 | ALMIDON MIDOVINIL |
| 4356.00 | MELOCOTON ANCESTRY EXTRA 425 | 10.57 | RAIDER CHOC. 3 X 58 GRS. | 4356.00 | MELOCOTON ANCESTRY EXTRA 425 |
| 4356.00 | MELOCOTON ANCESTRY EXTRA 850 | 17.18 | PRETTY TO ROW OIL 112 GRS. | 4356.00 | MELOCOTON ANCESTRY EXTRA 850 |

## Table 4 – Test of Cost Shocks as Cause

The 534-subsample is used. To construct a panel, the 29 months are divided into 5 half-year periods (but the last period is just 5 instead of 6 months). The specification for the upper half of the panel is:

$Log(AMPLIFICATION_t) = f(Log(AMPLIFICATION_{t-1})\ , PROMO\_COUNT_t\ , PROMO\_COUNT_{t+1})$ + fixed effects$_t$ ,

where $AMPLIFICATION_t$ is measured in the various forms listed in the top row, $PROMO\_COUNT_t$ is the number of months in each half-year period that has a trade promotion. In the lowr half, $PROMO\_COUNT_t$ is omitted. $PROMO\_DROP_t$ is also used as an alternate specification. It is the total percent price reduction from the regular cost of goods during the half-year period. Estimations are with fixed effects. Figures in brackets are standard errors. *** represents significance at 1%, ** at 5%, * at 10%. Only the variables of interest are shown in panels (b) and (c).

Panel (a) – SKU Level Estimations (*N* = 652)

| | Raw | | Seasonal | | Coint | | Amplitude | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Log(*AMPLI-* | -0.309 | -0.306 | -0.256 | -0.251 | -0.328 | -0.324 | -0.235 | -0.233 |
| *FICATION)*$_{t-1}$ | (0.054)*** | (0.054)*** | (0.054)*** | (0.055)*** | (0.049)*** | (0.049)*** | (0.049)*** | (0.049)*** |
| $PROMO\_COUNT_t$ | -0.048 | | -0.023 | | 0.198 | | -0.012 | |
| | (0.048) | | (0.027) | | (0.087)** | | (0.019) | |
| $PROMO\_COUNT_{t+1}$ | 0.079 | | 0.059 | | 0.043 | | 0.023 | |
| | (0.042)* | | (0.024)** | | (0.076) | | (0.017) | |
| $PROMO\_DROP_t$ | | -0.337 | | -0.102 | | 0.374 | | -0.194 |
| | | (0.526) | | (0.298) | | (0.946) | | (0.206) |
| $PROMO\_DROP_{t+1}$ | | -0.022 | | -0.009 | | 1.484 | | 0.081 |
| | | (0.544) | | (0.309) | | (0.977) | | (0.213) |
| Intercept | 1.497 | 1.492 | 0.764 | 0.799 | -0.488 | 0.121 | 0.335 | 0.337 |
| | (0.147)*** | (0.096)*** | (0.082)*** | (0.055)*** | (0.241)** | (0.132) | (0.054)*** | (0.033)*** |
| *R-squared* | 10.1% | 8.9% | 8.0% | 6.0% | 13.3% | 12.5% | 7.1% | 6.9% |
| *p value* | 0 | 0 | 0 | 0.0001 | 0 | 0 | 0 | 0 |
| Log(*AMPLI-* | -0.310 | -0.304 | -0.258 | -0.250 | -0.329 | -0.323 | -0.237 | -0.233 |
| *FICATION)*$_{t-1}$ | (0.054)*** | (0.054)*** | (0.054)*** | (0.055)*** | (0.049)*** | (0.049)*** | (0.049)*** | (0.049)*** |
| $PROMO\_COUNT_{t+1}$ | 0.083 | | 0.061 | | 0.026 | | 0.024 | |
| | (0.042)** | | (0.024)** | | (0.076) | | (0.016) | |
| $PROMO\_DROP_{+1}$ | | 0.146 | | 0.042 | | 1.301 | | 0.177 |
| | | (0.477) | | (0.270) | | (0.858) | | (0.187) |
| Intercept | 1.389 | 1.525 | 0.714 | 0.809 | -0.041 | 0.083 | 0.310 | 0.357 |
| | (0.099)*** | (0.081)*** | (0.056)*** | (0.047)*** | (0.141) | (0.091) | (0.035)*** | (0.026)*** |
| *R-squared* | 9.8% | 8.8% | 7.8% | 5.9% | 11.9% | 12.5% | 7.0% | 6.7% |
| *p value* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Panel (b) – Category Level Estimations (*N* = 204)

| | Raw | | Seasonal | | Coint | | Amplitude | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| *PROMO_COUNTt+1* | 0.000 | 0.146 | 0.019 | 0.042 | 0.019 | 1.301 | 0.003 | 0.177 |
| | (0.012) | (0.477) | (0.012) | (0.270) | (0.031) | (0.858) | (0.007) | (0.187) |
| *PROMO_DROP+1* | | 0.066 | | 0.062 | | 0.205 | | 0.056 |
| | | (0.090) | | (0.092) | | (0.230) | | (0.052) |
| *R-squared* | 12.6% | 12.9% | 11.8% | 10.4% | 6.6% | 7.0% | 12.3% | 13.0% |
| *p value* | 0.0002 | 0.0002 | 0.0004 | 0.0011 | 0.0137 | 0.0111 | 0.0003 | 0.0002 |

Panel (c) – Category Group Level Estimations (*N* = 91)

| | Raw | | Seasonal | | Coint | | Amplitude | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| *PROMO_COUNTt+1* | -0.004 | | 0.009 | | 0.008 | | 0.005 | |
| | (0.011) | | (0.013) | | (0.027) | | (0.006) | |
| *PROMO_DROP+1* | | 0.014 | | 0.122 | | 0.090 | | 0.009 |
| | | (0.076) | | (0.089) | | (0.192) | | (0.043) |
| *R-squared* | 26.7% | 26.5% | 4.9% | 7.2% | 5.9% | 6.1% | 18.3% | 17.5% |
| *p value* | 0.0003 | 0.0003 | 0.2624 | 0.1365 | 0.2021 | 0.1894 | 0.0047 | 0.0061 |

## Table 5 – Test of Demand Correlation as a Cause

### Panel (a) – SKU Level Estimations

To construct a panel, the 29 months are divided into 5 half-year periods (but the last period is just 5 instead of 6 months). The generic specification is:

$$Log(AMPLIFICATION_t) = f(Log(AMPLIFICATION_{t-1}),\ log(AR1_t)\text{ or }log(Var\ X_t)) + \text{fixed effects}_t\ ,$$

where $AMPLIFICATION_t$ is measured in the various forms listed in the top row, $AR1_t$ is the serial correlation obtained from regressions of $SALES_{it}$ on $SALES_{i,t-1}$, and $X_t$ is mark up multiplied by the sum of inventory, sales delivered, and sales unfulfilled. Estimations are with fixed effects. Figures in brackets are standard errors. *** represents significance at 1%, ** at 5%, * at 10%.

| | Raw | | Seasonal | | Cointegration | | Amplitude | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| **Full sample** | | | | | | | | |
| $L(AMP_{t-1})$ | -0.313 | -0.305 | -0.283 | -0.275 | -0.317 | -0.315 | -0.343 | -0.336 |
| | (0.016)*** | (0.016)*** | (0.016)*** | (0.016)*** | (0.017)*** | (0.018)*** | (0.015)*** | (0.015)*** |
| Log(*AR1*) | -0.012 | | -0.001 | | -0.014 | | -0.002 | |
| | (0.006)** | | (0.003) | | (0.012) | | (0.002) | |
| Log(Var*X*) | | 0.055 | | 0.009 | | 0.077 | | 0.045 |
| | | (0.013)*** | | (0.007) | | (0.026)*** | | (0.005)*** |
| Intercept | 1.578 | 1.038 | 0.867 | 0.780 | 0.031 | -0.729 | 0.407 | -0.036 |
| | (0.024)*** | (0.125)*** | (0.013)*** | (0.071)*** | (0.027) | (0.256)*** | (0.007)*** | (0.045) |
| *N* | 6786 | 6689 | 6786 | 6689 | 6780 | 6684 | 6791 | 6694 |
| *p value* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **534-sample** | | | | | | | | |
| $L(AMP_{t-1})$ | -0.293 | -0.268 | -0.253 | -0.212 | -0.410 | -0.283 | -0.229 | -0.235 |
| | (0.076)*** | (0.044)*** | (0.088)*** | (0.045)*** | (0.077)*** | (0.044)*** | (0.072)*** | (0.039)*** |
| Log(*AR1*) | -0.130 | | -0.016 | | -0.238 | | -0.029 | |
| | (0.077)* | | (0.051) | | (0.150) | | (0.032) | |
| Log(Var*X*) | | -0.009 | | 0.000 | | 0.018 | | 0.014 |
| | | (0.033) | | (0.019) | | (0.061) | | (0.012) |
| Intercept | 1.255 | 1.635 | 0.917 | 0.816 | -0.191 | -0.126 | 0.308 | 0.220 |
| | (0.155)*** | (0.344)*** | (0.100)*** | (0.201)*** | (0.223) | (0.625) | (0.054)*** | (0.129)* |
| *N* | 357 | 863 | 357 | 863 | 357 | 863 | 357 | 864 |
| *p value* | 0.0002 | 0 | 0.0165 | 0 | 0 | 0 | 0.005 | 0 |

### Panel (b) – Category Level Estimations

| | Raw | | Seasonal | | Cointegration | | Amplitude | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| **Full sample** | | | | | | | | |
| Log(*AR1*) | -0.015 | | 0.001 | | -0.013 | | 0.000 | |
| | (0.014) | | (0.013) | | (0.046) | | (0.008) | |
| Log(Var *X*) | | 0.076 | | 0.013 | | 0.126 | | 0.042 |
| | | (0.018)*** | | (0.017) | | (0.061)** | | (0.011)*** |
| *N* | 1445 | 1443 | 1445 | 1443 | 1445 | 1443 | 1445 | 1443 |
| *p value* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **534-sample** | | | | | | | | |
| Log(*AR1*) | 0.004 | | 0.379 | | 0.287 | | 0.037 | |
| | (0.118) | | (0.103)*** | | (0.253) | | (0.062) | |
| Log(Var *X*) | | 0.198 | | 0.050 | | 0.120 | | 0.084 |
| | | (0.034)*** | | (0.033) | | (0.081) | | (0.018)*** |
| *N* | 277 | 277 | 276 | 276 | 277 | 277 | 277 | 277 |
| *p value* | 0.0639 | 0 | 0 | 0.0022 | 0.0101 | 0.0065 | 0.0004 | 0 |

## Table 6 – Test of Batching as a Cause

Two estimation methods are used. The first is to use instrumental variables: for SKU (ratio of weight to markup) and for retailer (private label dummy and $R$-squared of seasonal components of regression of sales delivered on month dummies for each SKU). There is no instrumental variable for batching by supplier, so the regression is just OLS. The full subsample is used for this cross-sectional regression. The second method is fixed effects on a panel. The 534-subsample is used. To construct a panel, the 29 months are divided into 5 half-year periods (but the last period is just 5 instead of 6 months). The generic specification is:

$$Log(AMPLIFICATION_t) = f(Log(AMPLIFICATION_{t-1}),\ BATCH_t) + \text{fixed effects}_t,$$

where $AMPLIFICATION_t$ is measured in the various forms listed in the top row, $BATCH_t$ is measured in three ways: by SKU (proxied by average months between orders over the half-year periods), supplier (average months between orders for all SKUs from the supplier), and retailer ($R$-squared of seasonal components of regression of supplies received on month dummies for each SKU). The column for the stochastic amplification ratio does not have a fixed effects estimation because stochastic ratios cannot be estimated with the panel. This is because calculating stochastic ratios require regressions that are under-identified, with more variables (e.g., 12 monthly dummies) than observations per SKU (i.e., 6 months of data for each half-year period). Figures in brackets are standard errors. *** represents significance at 1%, ** at 5%, * at 10%.

Panel (a) – SKU Level Estimations

| | Raw | | Seasonal | | Stochastic | Cotin | | Amplitude | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) IV | (2) Fixed E. | (3) IV | (4) Fixed E. | (5) IV | (6) IV | (7) Fixed E. | (8) IV | (9) Fixed E. |
| **Batching by SKU** | | | | | | | | | |
| *L( $AMP_{t-1}$)* | | -0.282 | | -0.221 | | | -0.289 | | -0.248 |
| | | (0.043)*** | | (0.044)*** | | | (0.042)*** | | (0.037)*** |
| *BATCH-* | -0.188 | 0.175 | -0.039 | 0.074 | -0.305 | 0.214 | 0.231 | 0.029 | 0.087 |
| *SKU* | (0.252) | (0.054)*** | (0.102) | (0.032)** | (0.321) | (0.387) | (0.097)** | (0.096) | (0.020)*** |
| Intercept | 1.630 | 1.431 | 0.669 | 0.768 | 2.218 | -0.947 | -0.074 | 0.305 | 0.308 |
| | (1.040) | (0.074)*** | (0.421) | (0.043)*** | (1.326)* | (1.596) | (0.088) | (0.397) | (0.023)*** |
| *p value* | 0.4543 | 0 | 0.7016 | 0 | 0.3426 | 0.5796 | 0 | 0.7599 | 0 |
| *N* | 3180 | 879 | 3180 | 879 | 3180 | 3180 | 879 | 3180 | 880 |
| **Batching by Supplier** | | | | | | | | | |
| *L($AMP_{t-1}$)* | | -0.277 | | -0.219 | | | -0.283 | | -0.231 |
| | | (0.043)*** | | (0.044)*** | | | (0.042)*** | | (0.038)*** |
| *BATCH-* | 0.053 | 0.297 | 0.025 | 0.137 | 0.044 | 0.059 | 0.572 | 0.033 | 0.168 |
| *SUPPLIER* | (0.012)*** | (0.115)** | (0.006)*** | (0.067)** | (0.012)*** | (0.015)*** | (0.204)*** | (0.006)*** | (0.043)*** |
| Intercept | 0.812 | 1.480 | 0.487 | 0.787 | 0.932 | -0.125 | -0.036 | 0.403 | 0.325 |
| | (0.014)*** | (0.070)*** | (0.008)*** | (0.040)*** | (0.015)*** | (0.024)*** | (0.074) | (0.008)*** | (0.021)*** |
| *p value* | 0 | 0 | 0 | 0 | 0.0002 | 0.0001 | 0 | 0 | 0 |
| *N* | 3723 | 879 | 3723 | 879 | 3723 | 3723 | 879 | 3723 | 880 |
| **Batching by Retailer** | | | | | | | | | |
| *L( $AMP_{t-1}$)* | | -0.290 | | -0.228 | | | -0.288 | | -0.250 |
| | | (0.043)*** | | (0.044)*** | | | (0.042)*** | | (0.038)*** |
| *BATCH-* | 0.752 | - | 0.397 | - | 0.058 | 0.382 | - | 0.264 | - |
| *RETAILER* | (0.087)*** | | (0.052)*** | | (0.125) | (0.132)*** | | (0.052)*** | |
| Intercept | 0.574 | 1.548 | 0.359 | 0.818 | 0.952 | -0.214 | 0.066 | 0.333 | 0.361 |
| | (0.035)*** | (0.065)*** | (0.020)*** | (0.038)*** | (0.048)*** | (0.054)*** | (0.065) | (0.021)*** | (0.019)*** |
| *p value* | 0 | 0 | 0 | 0 | 0.6391 | 0 | 0 | 0 | 0 |
| *N* | 3723 | 879 | 3723 | 879 | 3723 | 3723 | 879 | 3723 | 880 |

Panel (b) - Category Level Estimations

| | Raw | | Seasonal | | Stochastic | Cotin | | Amplitude | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| | IV | Fixed E. | IV | Fixed E. | IV | IV | Fixed E. | IV | Fixed E. |
| **Batching by SKU** | | | | | | | | | |
| *$L( AMP_{t-1})$* | | -0.145 | | -0.209 | | | -0.191 | | -0.224 |
| | | (0.059)** | | (0.065)*** | | | (0.068)*** | | (0.054)*** |
| *BATCH-* | 0.043 | 0.113 | -0.010 | 0.005 | 0.004 | 0.027 | -0.171 | 0.126 | 0.069 |
| *SKU* | (0.101) | (0.065)* | (0.048) | (0.059) | (0.135) | (0.070) | (0.141) | (0.061)** | (0.034)** |
| Intercept | 0.527 | 0.641 | 0.516 | 0.753 | 0.620 | -0.051 | -0.131 | 0.085 | 0.278 |
| | (0.217)** | (0.062)*** | (0.102)*** | (0.064)*** | (0.292)** | (0.143) | (0.115) | (0.129) | (0.031)*** |
| *N* | 392 | 277 | 392 | 276 | 392 | 392 | 277 | 392 | 277 |
| *p value* | 0.6696 | 0.0146 | 0.8283 | 0.0065 | 0.9735 | 0.7037 | 0.0092 | 0.0404 | 0.0001 |
| **Batching by Retailer** | | | | | | | | | |
| *$L( AMP_{t-1})$* | | -0.134 | | -0.203 | | | -0.195 | | -0.215 |
| | | (0.059)** | | (0.065)*** | | | (0.069)*** | | (0.055)*** |
| *BATCH-* | 1.145 | 0.899 | 0.362 | 0.830 | 1.390 | 0.242 | 0.090 | 0.356 | 0.209 |
| *RETAILER* | (0.210)*** | (0.490)* | (0.140)** | (0.444)* | (0.225)*** | (0.276) | (1.063) | (0.130)*** | (0.259) |
| Intercept | 0.492 | 0.422 | 0.457 | 0.491 | 0.472 | -0.014 | -0.266 | 0.317 | 0.253 |
| | (0.026)*** | (0.163)** | (0.020)*** | (0.151)*** | (0.027)*** | (0.034) | (0.343) | (0.021)*** | (0.086)*** |
| *N* | 408 | 277 | 408 | 276 | 408 | 408 | 277 | 408 | 277 |
| *p value* | 0 | 0.0121 | 0.0102 | 0.0012 | 0 | 0.3822 | 0.0192 | 0.0063 | 0.0003 |

## Table 7 – Test of Gaming as a Cause

Panel (a) – Sherry Labor Strike as a Natural Experiment - Raw Amplification Ratio

The 29 months are divided into five 6-month periods (except the last, which is 5 months), indexed below. The sherry labor strike starts November 1990, lasting 59 days. This is the end of period 2. This dataset is at the SKU level, on the full sample. In the rightmost two parts (SKUs in the same category group or same category), sherry SKUs are excluded from the tallies. The following data is for the raw amplification ratio, which is winsorized at 1% and 99%.

| | Sherry SKUs | | | All SKUs | | | All SKUs except sherry | | | SKUs in same category group as sherry | | | SKUs in same category as sherry | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Time* | *N* | *Mean* | *S.D.* | *N* | *Mean* | *S.D.* | *N* | *Mean* | *S.D.* | *N* | *Mean* | *S.D.* | *N* | *Mean* | *S.D.* |
| 1 | 13 | 16.41 | 28.03 | 3745 | 48.16 | 47.41 | 3732 | 48.27 | 47.43 | 253 | 51.84 | 47.10 | 30 | 65.79 | 45.86 |
| 2 | 13 | 10.03 | 27.11 | 3745 | 41.97 | 46.89 | 3732 | 42.08 | 46.91 | 253 | 37.46 | 46.40 | 30 | 55.03 | 48.98 |
| 3 | 13 | 14.43 | 27.56 | 3745 | 31.47 | 43.34 | 3732 | 31.52 | 43.38 | 253 | 41.38 | 45.53 | 30 | 42.34 | 45.61 |
| 4 | 13 | 11.76 | 27.46 | 3745 | 32.82 | 43.03 | 3732 | 32.89 | 43.06 | 253 | 26.45 | 40.25 | 30 | 22.59 | 36.40 |
| 5 | 13 | 27.95 | 42.11 | 3745 | 30.03 | 42.43 | 3732 | 30.04 | 42.44 | 253 | 21.81 | 36.93 | 30 | 27.61 | 42.20 |

Panel (b) – Sherry Labor Strike as a Natural Experiment - Estimations

The dataset is on the full sample. The specification is: $Log(AMPLIFICATION_t) = f\,(Log(AMPLIFICATION_{t-1}, EVENT?_t\,,\ SHERRY?_t\,,\ EVENT?_t \times SHERRY?_t)$, where $EVENT?_t$ Is a dummy set to 1 if the time is one period before the supposed shortage kicks in (which may be during or after the strike) and $SHERRY?_t$ Is 1 if the SKU is a sherry, or if the category or category group contains a sherry SKU. All estimations are done with robust standad errors and clustered around SKU, category, or category group, respectively. The $SHERRY?_t$ variable drops out from the fixed effects estimations since it is time-invariant.

| | Sherry SKUs | | | | SKUs in same category as sherry | | | | SKUs in same category group as sherry | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Raw | Seasonal | Coint | Amplitude | Raw | Seasonal | Coint | Amplitude | Raw | Seasonal | Coint | Amplitude |
| *L( $AMP_{t-1}$)* | 0.000 | 0.000 | -0.032 | 0.045 | 0.266 | 0.274 | 0.226 | 0.325 | 0.295 | 0.304 | 0.254 | 0.356 |
| | (0.008) | (0.008) | (0.008)*** | (0.007)*** | (0.007)*** | (0.007)*** | (0.007)*** | (0.007)*** | (0.007)*** | (0.007)*** | (0.007)*** | (0.007)*** |
| *EVENT?* | 0.073 | 0.034 | 0.119 | 0.052 | 0.139 | 0.066 | 0.204 | 0.087 | 0.189 | 0.091 | 0.245 | 0.107 |
| at time=2 | (0.032)** | (0.017)** | (0.043)*** | (0.014)*** | (0.033)*** | (0.018)*** | (0.045)*** | (0.015)*** | (0.035)*** | (0.019)*** | (0.047)*** | (0.016)*** |
| Interaction | **-1.454** | **-0.819** | **-1.263** | **-0.526** | **-0.789** | **-0.382** | **-0.511** | **-0.295** | **-0.735** | **-0.363** | **-0.530** | **-0.276** |
| | **(0.535)*** ** | **(0.295)*** ** | **(0.730)*** | **(0.239)** ** | **(0.311)** ** | **(0.171)** ** | **(0.422)** | **(0.141)** ** | **(0.130)*** ** | **(0.071)*** ** | **(0.176)*** ** | **(0.059)*** ** |
| Intercept | 2.648 | 1.474 | 1.923 | 0.996 | 1.832 | 1.017 | 1.312 | 0.633 | 1.744 | 0.968 | 1.245 | 0.593 |
| | (0.028)*** | (0.016)*** | (0.027)*** | (0.012)*** | (0.027)*** | (0.015)*** | (0.027)*** | (0.011)*** | (0.027)*** | (0.015)*** | (0.027)*** | (0.011)*** |
| *R-sq* | 0.0071 | 0.0118 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *p value* | 0.1% | 0.2% | 78.6% | 85.1% | 90.7% | 90.8% | 90.2% | 93.4% | 87.3% | 88.4% | 89.4% | 93.2% |
| *N* | 18725 | 18725 | 18725 | 18725 | 18725 | 18725 | 18725 | 18725 | 18725 | 18725 | 18725 | 18725 |
| Interaction, | **0.160** | **0.192** | **-0.165** | **0.020** | **0.204** | **0.122** | **-0.121** | **0.031** | **0.179** | **0.087** | **0.133** | **0.010** |
| event=1 | **(0.533)** | **(0.294)** | **(0.728)** | **(0.239)** | **(0.310)** | **(0.171)** | **(0.421)** | **(0.140)** | **(0.129)** | **(0.071)** | **(0.175)** | **(0.059)** |
| Interaction, | **0.261** | **0.116** | **-0.667** | **-0.057** | **0.370** | **0.167** | **-0.077** | **0.126** | **0.786** | **0.424** | **0.679** | **0.342** |
| event=3 | **(0.534)** | **(0.294)** | **(0.730)** | **(0.239)** | **(0.311)** | **(0.171)** | **(0.422)** | **(0.141)** | **(0.130)*** ** | **(0.071)*** ** | **(0.176)*** ** | **(0.059)*** ** |

Continued…

(c) SKU Level Estimations

The following specification is used, with two-stage-least-squares fixed effects on a panel of the 534-subsample:

$$Log(AMPLIFICATION_t) = f(Log(AMPLIFICATION_{t-1}, GAMING_{it}) + \text{fixed effects}_t,$$

where $AMPLIFICATION_t$ is measured in the various forms listed in the top row, $GAMING_t$ is measured in various ways, depicted in the leftmost column. Each $GAMING_{it}$ variable in the structural equation is instrumented with all other gaming variables in the leftmost column. The *p*-value of the $\chi(4)$ statistic for a test of over-identifying restrictions, and is obtained by regressing the residual of the structural equation on all exogenous variables, and then taking the *N.R*-squared at $\chi(4)$, since there are four remaining exogenous instruments. The stochastic amplification ratio is not used because stochastic ratios cannot be estimated with the panel. This is because calculating stochastic ratios require regressions that are under-identified, with more variables (e.g., 12 monthly dummies) than observations per SKU (i.e., 6 months of data for each half-year period). Figures in brackets are standard errors. *** represents significance at 1%, ** at 5%, * at 10%.

| | Raw | | Seasonal | | Coint | | Amplitude | |
|---|---|---|---|---|---|---|---|---|
| | IV | *p*, $\chi(4)$ | IV | *p*, $\chi(4)$ | IV | *p*, $\chi(4)$ | IV | *p*, $\chi(4)$ |
| 1/brands in category | -0.923 (0.739) | 0.000 | -0.262 (0.431) | 0.063 | 0.317 (1.324) | 0.071 | -0.379 (0.279) | 0.000 |
| - returns to suppliers | 1.687 (1.333) | 0.608 | 0.695 (0.778) | 0.520 | 1.679 (2.398) | 0.082 | 0.548 (0.503) | 0.057 |
| - returns fr customers | 0.399 (0.728) | 0.546 | 0.184 (0.424) | 0.837 | -0.559 (1.305) | 0.216 | 0.138 (0.275) | 0.071 |
| Total stock out months | -0.005 (0.039) | 0.619 | 0.000 (0.023) | 0.736 | 0.014 (0.069) | 0.332 | 0.000 (0.015) | 0.033 |
| Abnormal sales | 0.016 (0.055) | 0.345 | 0.002 (0.032) | 0.681 | -0.105 (0.098) | 0.036 | -0.002 (0.021) | 0.045 |

(d) Category and Category Group Levels Estimations

| | Raw | | Seasonal | | Coint | | Amplitude | |
|---|---|---|---|---|---|---|---|---|
| | Cat | Cat grp | Cat | Cat grp | Cat | Cat grp | Cat | Cat grp |
| 1/brands in category | -0.959 (0.401)** | -0.307 (0.849) | 0.059 (0.369) | 0.270 (0.896) | 0.083 (0.872) | -1.937 (1.700) | -0.089 (0.213) | -0.010 (0.461) |
| - returns to suppliers | 1.187 (1.508) | -1.375 (3.495) | 0.843 (1.365) | -2.560 (3.680) | -3.064 (3.179) | 7.239 (7.040) | 1.662 (0.778)** | -1.534 (1.871) |
| - returns fr customers | -0.414 (0.451) | 0.924 (1.109) | -0.183 (0.410) | -1.506 (1.157) | -0.518 (0.962) | 0.748 (2.225) | -0.036 (0.236) | 0.110 (0.595) |
| Total stock out months | -0.006 (0.009) | -0.006 (0.006) | -0.005 (0.008) | -0.009 (0.007) | -0.015 (0.019) | -0.007 (0.013) | -0.006 (0.005) | -0.006 (0.003)* |
| Abnormal sales | 0.015 (0.030) | 0.068 (0.067) | 0.026 (0.027) | -0.005 (0.072) | 0.084 (0.064) | 0.280 (0.136)** | -0.010 (0.016) | 0.006 (0.037) |

## Table 8 – Test of Behavioral Under-reaction and Over-reaction as Causes

The estimations are with robust standard errors and clustered. Figures in brackets are standard errors. *** represents significance at 1%, ** at 5%, * at 10%.

Panel (a) – SKU Level Estimation of Under-reaction

The odd-numbered models use the full sample, but limit to observations in which the absolute value of *AR1*, the demand correlation coefficient for sales, is smaller than 0.1. The models are estimated with the following in a cross-section:

$$Log(AMPLIFICATION_i) = \beta_0 + \beta_1.LAG_{it} + \varepsilon_t \text{ ,}$$

where $LAG_i$ is a dummy for SKUs which do have replenishment lags, and $\varepsilon_t$ is i.i.d. No lag of amplification is needed since $LAG_i$ is exogenous. In the even-numbered models, all observations are used, but $log(AR1_i)$ is used as a control.

| | Raw | | Seasonal | | Stochastic | | Coint | | Amplitude | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| $LAG_i$ | 0.261 | 0.292 | 0.098 | 0.128 | 0.265 | 0.313 | 0.049 | 0.041 | 0.118 | 0.158 |
| | (0.041)*** | (0.039)*** | (0.025)*** | (0.026)*** | (0.045)*** | (0.043)*** | (0.086) | (0.085) | (0.026)*** | (0.026)*** |
| $Log(AR1_i)$ | | -0.003 | | 0.053 | | 0.042 | | -0.096 | | 0.063 |
| | | (0.012) | | (0.008)*** | | (0.014)*** | | (0.022)*** | | (0.008)*** |
| Intercept | 0.617 | 0.576 | 0.419 | 0.458 | 0..724 | 0.722 | -0.113 | -0.232 | 0.323 | 0.367 |
| | (0.039)*** | (0.038)*** | (0.024)*** | (0.025)*** | (0.043)*** | (0.042)*** | (0.083) | (0.087)*** | (0.025)*** | (0.026)*** |
| *N* | 3723 | 3310 | 3723 | 3310 | 3723 | 3310 | 3723 | 3310 | 3723 | 3310 |
| *p* value | 0.000 | 0 | 0.000 | 0 | 0.000 | 0 | 0.569 | 0.0001 | 0.000 | 0 |

Panel (b) – Category Level Estimation of Under-reaction

| | Raw | | Seasonal | | Stochastic | | Coint | | Amplitude | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| $LAG_i$ | 0.190 | 0.181 | 0.117 | 0.126 | 0.194 | 0.183 | 0.011 | 0.028 | 0.085 | 0.090 |
| | (0.084)** | (0.088)** | (0.078) | (0.073)* | (0.085)** | (0.089)** | (0.053) | (0.049) | (0.086) | (0.085) |
| $Log(AR1_i)$ | | -0.106 | | 0.101 | | -0.123 | | 0.234 | | 0.043 |
| | | (0.080) | | (0.047)** | | (0.082) | | (0.197) | | (0.034) |
| Intercept | 0.453 | 0.379 | 0.390 | 0.460 | 0.459 | 0.373 | 0.006 | 0.169 | 0.279 | 0.309 |
| | (0.081)*** | (0.102)*** | (0.076)*** | (0.079)*** | (0.082)*** | (0.104)*** | (0.042) | (0.144) | (0.085)*** | (0.087)*** |
| *N* | 408 | 405 | 408 | 405 | 408 | | 408 | 405 | 408 | 405 |
| *p* value | 0.0248 | 0.0552 | 0.1346 | 0.0226 | 0.0234 | | 0.8279 | 0.0436 | 0.3258 | 0.2496 |

Panel (c) – SKU Level Estimation of Over-reaction

The odd and even-numbered models are estimated as in panel (a), but the baseline specification is $Log(AMPLIFICATION_i) = \beta_0 + \beta_1.AR1\_SUPPLIERS_i + \varepsilon_t$ .

| | Raw | | Seasonal | | Stochastic | | Coint | | Amplitude | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| *AR1_* | -0.726 | -0.879 | -0.327 | -0.505 | -0.794 | -1.036 | 0.134 | 0.370 | -0.328 | -0.541 |
| $SUPPLIERS_i$ | (0.138)*** | (0.192)*** | (0.062)*** | (0.102)*** | (0.147)*** | (0.215)*** | (0.055)** | (0.066)*** | (0.066)*** | (0.112)*** |

## Table 9 – Test of Coordination Risk as a Cause

Figures in brackets are standard errors. *** represents significance at 1%, ** at 5%, * at 10%.

### Panel (a) – SKU Level Estimation

The first specification is a an OLS cross-section estimation of the following:

$$Log(AMPLIFICATION_t) = f(\text{Log}(SUPPLIER\_VALUE_i), SUPPLIER\_SKUS_i, SUPPLIER\_WEIGHT\_SD_i) + \varepsilon_t ,$$

where $SUPPLIER\_VALUE_i$ is the total value of goods delivered by the supplier of SKU $i$ to the retailer over the full 29 months of the cross-section, $SUPPLIER\_SKUS_i$ the number of SKUs from the supplier of SKU $i$, and $SUPPLIER\_WEIGHT\_SD_i$ the standard deviation of the weights of the SKUs from the supplier of SKU $i$. The estimation is with robust standard errors and clustered at the SKU or category levels in panels (a) and (b), respectively. The second is a panel fixed effects estimation of:

$$Log(AMPLIFICATION_t) = f(Log(AMPLIFICATION_{t-1}), MTHS\_SINCE\ LAST\_SALE_{it}) + \text{fixed effects}_t ,$$

where $MTHS\_SINCE\ LAST\_SALE_{it}$ is the number of months since the last sale as of time $t$ for SKU $i$. The 534-subsample is used in the construction of the panel. The 29 months are divided into five 6-month periods (except the last, which is 5 months). To avoid truncation bias, the first 6-month period is not used.

| | Raw | | Seasonal | | Stochastic | Coint | | Amplitude | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) Fixed eff. | (3) OLS | (4) Fixed eff. | (5) OLS | (6) OLS | (7) Fixed eff. | (8) OLS | (9) Fixed eff. |
| Log(*SUPPLIER_ VALUE$_i$*) | 0.070 | | 0.046 | | 0.067 | -0.028 | | 0.062 | |
| | (0.012)*** | | (0.006)*** | | (0.012)*** | (0.020) | | (0.006)*** | |
| *SUPPLIER_ SKUS$_i$* | 0.002 | | 0.002 | | 0.002 | -0.005 | | 0.002 | |
| | (0.001)*** | | (0.000)*** | | (0.001)*** | (0.001)*** | | (0.000)*** | |
| *SUPPLIER_ WEIGHT_SD$_i$* | 0.000 | | 0.000 | | 0.000 | 0.002 | | 0.000 | |
| | (0.001) | | (0.000) | | (0.001) | (0.002) | | (0.001) | |
| Log(*AMPLIFI- CATION$_{t,t-1}$*) | | -0.293 | | -0.232 | | | -0.280 | | -0.225 |
| | | (0.044)*** | | (0.045)*** | | | (0.042)*** | | (0.039)*** |
| *MTHS_SINCE_ LAST_SALE$_{it}$* | | -0.040 | | -0.033 | | | 0.366 | | 0.125 |
| | | (0.138) | | (0.080) | | | (0.241) | | (0.052)** |
| Intercept | 2.027 | 1.555 | 1.252 | 0.824 | | -0.452 | 0.031 | 1.459 | 0.341 |
| | (0.198)*** | (0.070)*** | (0.108)*** | (0.040)*** | | (0.346) | (0.069) | (0.110)*** | (0.021)*** |
| *N* | 3693 | 879 | 3693 | 879 | 3693 | 3693 | 879 | 3693 | 880 |
| *p* value | 0 | 0 | 0 | 0 | 0 | 0.0039 | 0 | 0 | 0 |

### Panel (b) – Category Level Estimation

The specification is the same as the second model in panel (a).

| | Raw | Seasonal | Coint | Amplitude |
|---|---|---|---|---|
| Log(*AMPLIFI- CATION$_{t,t-1}$*) | -0.132 | -0.211 | -0.194 | -0.216 |
| | (0.059)** | (0.065)*** | (0.069)*** | (0.055)*** |
| *MTHS_SINCE LAST_SALE$_{it}$* | 0.298 | -0.060 | 0.142 | 0.136 |
| | (0.167)* | (0.153) | (0.360) | (0.087) |
| Intercept | 0.654 | 0.767 | -0.261 | 0.297 |
| | (0.057)*** | (0.059)*** | (0.094)*** | (0.027)*** |
| *N* | 277 | 276 | 277 | 277 |
| *p* value | 0.0131 | 0.0061 | 0.0178 | 0.0001 |

## Table 10 – Test of All Causes

Two estimations are done for each amplification ratio in each panel. The first is an OLS, on the full sample cross-section. The measures are: *PROMO_COUNT$_t$* for cost shock, *log*(*AR1$_t$* ) for demand correlation, *BATCH_SKU$_i$* , *BATCH_SUPPLIER$_i$*, and *BATCH_RETAILER$_i$* for batching, reciprocal of *NUM_OF_BRANDS_IN_CAT$_i$* for gaming, *LAG$_i$* for under-reaction and Log(1/*SUPPLIER_ VALUE$_i$*) and *SUPPLIER_ SKUS$_i$* for coordination. The estimation is with robust standard errors and clustered at the SKU or category levels in panels (a) and (b), respectively. The second method is a fixed effects estimation using the 534-subsample panel, in which the 29 months are divided into five 6-month periods (except the last, which is 5 months). All variables are indexed over time and a lagged dependant variable is added. The *R*-squared is the adjusted one for OLS and the within one for fixed effects. Figures in brackets are standard errors. *** represents significance at 1%, ** at 5%, * at 10%.

Panel (a) – SKU Level Estimation

| | Raw | | Season. | | Stoch. | Coint. | | Amplitude | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) Fixed e. | (3) OLS | (4) Fixed e. | (5) OLS | (7) OLS | (8) Fixed e. | (9) OLS | (10)Fixed e. |
| L(*AMPLIFICATION$_{t,t-1}$*) | | -0.360 | | -0.293 | | | -0.328 | | -0.412 |
| | | (0.034)*** | | (0.037)*** | | | (0.037)*** | | (0.034)*** |
| Cost shock | 0.002 | 0.089 | -0.003 | 0.056 | 0.002 | 0.020 | 0.077 | 0.001 | 0.039 |
| | (0.002) | (0.032)*** | (0.001)** | (0.020)*** | (0.003) | (0.004)*** | (0.066) | (0.001) | (0.013)*** |
| Demand correlation | 0.014 | -0.011 | 0.057 | 0.049 | 0.040 | -0.042 | 0.033 | 0.067 | 0.011 |
| | (0.015) | (0.029) | (0.009)*** | (0.018)*** | (0.017)** | (0.026) | (0.060) | (0.010)*** | (0.012) |
| Batching within SKUs | 0.042 | 0.094 | 0.019 | 0.038 | 0.038 | 0.014 | 0.005 | 0.030 | 0.043 |
| | (0.003)*** | (0.051)* | (0.002)*** | (0.032) | (0.004)*** | (0.007)* | (0.107) | (0.002)*** | (0.021)** |
| Batching to supplier | -0.010 | 0.378 | -0.007 | 0.298 | -0.016 | 0.046 | 0.581 | -0.010 | 0.220 |
| | (0.009) | (0.154)** | (0.005) | (0.096)*** | (0.010) | (0.014)*** | (0.319)* | (0.005)* | (0.062)*** |
| Batching of retailers | 0.808 | (dropped) | 0.518 | (dropped) | 0.133 | 0.269 | (dropped) | 0.400 | (dropped) |
| | (0.094)*** | (0.000)*** | (0.056)*** | (0.000)*** | (0.141) | (0.139)* | (0.000)*** | (0.055)*** | (0.000)*** |
| Gaming | -0.042 | -1.065 | -0.025 | -0.600 | -0.033 | 0.083 | -2.710 | -0.024 | -0.435 |
| | (0.019)** | (0.474)** | (0.011)** | (0.296)** | (0.021) | (0.032)*** | (0.985)*** | (0.012)** | (0.190)** |
| Under-reaction | 0.231 | | 0.101 | | 0.276 | -0.019 | | 0.117 | |
| | (0.038)*** | | (0.026)*** | | (0.044)*** | (0.090) | | (0.026)*** | |
| Coordination – | 0.002 | | 0.001 | | 0.002 | -0.004 | | 0.002 | |
| *SUPPLIER_ SKUS$_i$* | (0.001)*** | | (0.000)*** | | (0.001)** | (0.001)*** | | (0.000)*** | |
| Coordination – | 0.030 | | 0.016 | | 0.033 | -0.076 | | 0.033 | |
| L(1/*SUPPLIER_ VALUE$_i$*) | (0.011)*** | | (0.006)*** | | (0.013)*** | (0.021)*** | | (0.006)*** | |
| Intercept | 0.655 | 1.472 | 0.491 | 0.971 | 1.079 | -1.654 | 0.411 | 0.681 | 0.430 |
| | (0.182)*** | (0.124)*** | (0.108)*** | (0.077)*** | (0.217)*** | (0.359)*** | (0.237)* | (0.109)*** | (0.048)*** |
| *N* | 3310 | 2306 | 3310 | 2306 | 3310 | 3310 | 2306 | 3310 | 2307 |
| *p* value | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *R*-squared | 13.7% | 18.7% | 12.2% | 14.2% | 7.1% | 2.4% | 12.4% | 18.0% | 24.1% |

Panel (b) – Category Level Estimation

The specifications are the same as the ones for the SKU level estimations, except that some variables (the ones not shown) are dropped because they do not exist at the category level.

| | Raw | | Season. | | Stoch. | Coint. | | Amplitude | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) Fixed eff. | (3) OLS | (4) Fixed eff. | (5) OLS | (7) OLS | (8) Fixed eff. | (9) OLS | (10) Fixed eff. |
| Log(*AMPLIFICATION*$_{t,t-1}$) | | -0.402 | | -0.297 | | | -0.369 | | -0.389 |
| | | (0.029)*** | | (0.033)*** | | | (0.031)*** | | (0.031)*** |
| Cost shock | -0.001 | -0.006 | -0.001 | -0.002 | -0.001 | 0.001 | -0.006 | -0.001 | -0.004 |
| | (0.000)*** | (0.005) | (0.000)*** | (0.005) | (0.000)*** | (0.000) | (0.008) | (0.000)*** | (0.003) |
| Demand correlation | -0.025 | 0.111 | 0.135 | 0.206 | -0.034 | 0.295 | 0.186 | 0.079 | 0.101 |
| | (0.080) | (0.030)*** | (0.051)*** | (0.029)*** | (0.082) | (0.202) | (0.051)*** | (0.037)** | (0.019)*** |
| Batching within SKUs | 0.052 | -0.003 | 0.017 | -0.027 | 0.043 | 0.063 | -0.340 | 0.038 | 0.005 |
| | (0.014)*** | (0.043) | (0.008)** | (0.041) | (0.015)*** | (0.020)*** | (0.073)*** | (0.009)*** | (0.026) |
| Batching of retailers | 0.948 | 1.459 | 0.512 | 1.797 | 1.200 | 0.308 | 0.804 | 0.237 | 0.807 |
| | (0.258)*** | (0.277)*** | (0.185)*** | (0.266)*** | (0.295)*** | (0.411) | (0.466)* | (0.159) | (0.171)*** |
| Gaming | -0.070 | -0.109 | -0.042 | 0.117 | -0.061 | -0.012 | 0.522 | -0.042 | 0.093 |
| | (0.027)*** | (0.144) | (0.016)** | (0.137) | (0.026)** | (0.033) | (0.241)** | (0.018)** | (0.089) |
| Intercept | 0.486 | 0.669 | 0.567 | 0.601 | 0.473 | 0.047 | -0.161 | 0.389 | 0.321 |
| | (0.056)*** | (0.100)*** | (0.039)*** | (0.097)*** | (0.056)*** | (0.107) | (0.167) | (0.032)*** | (0.062)*** |
| *N* | 405 | 1034 | 405 | 1034 | 405 | 405 | 1034 | 405 | 1034 |
| *p* value | 0 | 0 | 0 | 0 | 0 | 0.0196 | 0 | 0 | 0 |
| *R*-squared | 22.7% | 28.6% | 11.6% | 22.7% | 23.3% | 12.9% | 26.6% | 14.2% | 26.2% |