ARTICLE IN PRESS
Journal of Econometrics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Estimation of endogenously sampled time series: The case of commodity price speculation in the steel market

George Hall [a], John Rust [b,*,1]

[a] Brandeis University, United States of America
[b] Georgetown University, United States of America

ARTICLE INFO

Article history:
Available online xxxx

JEL classification:
C13-15

Keywords:
Endogenous sampling
Markov processes
Simulation estimation
Method of simulated moments (MSM)
Commodity price speculation
Inventory investment
(S, s) strategies

ABSTRACT

We consider the problem of estimating a Markov process that is *endogenously sampled.* We observe a discrete-time Markov process $\{p_t\}$ at a set of random times $\{t_1, \ldots, t_n\}$ that depend on the outcome of a probabilistic sampling rule that depends on the state of the process and other observed covariates x_t. We focus on a particular example where p_t is the daily wholesale price of a standardized steel product. The endogenous sampling problem arises from the fact that we only observe p_t on the days the firm purchases steel. We show how to solve this problem under two different assumptions about firm behavior: (1) *optimality:* the timing of steel purchases is governed by an optimal purchasing strategy that maximizes expected discounted profits, and (2) *potential suboptimality:* we allow the firm to use any randomized, Markovian purchasing strategy. In the latter case, the estimation problem becomes semi-parametric and we use the method of sieves to estimate a flexible parametric approximation to the firm's purchasing behavior that best fits the data without imposing optimality. We show how estimation of this model becomes tractable under either of these assumptions using the method of simulated moments (MSM). We simulate realizations of wholesale steel prices and sample them in the same way as they are sampled in the actual data, i.e. only on days where purchases occur. We use the MSM estimator to estimate a truncated lognormal AR(1) model of the wholesale price processes for particular types of steel plate and test and reject the assumption that the firm is behaving optimally.
© 2020 Elsevier B.V. All rights reserved.

1. Introduction

This paper studies the problem of estimating a Markov process that is *endogenously sampled.* Our objective is to estimate the parameters of a discrete-time Markov process $\{p_t\}$ that is observed only at a subset of random times $\{t_1, \ldots, t_n\}$ that

* Correspondence to: Department of Economics, Georgetown University, Washington, DC 20057, United States of America.
E-mail address: jrust@editorialexpress.com (J. Rust).
URL: http://editorialexpress.com/jrust (J. Rust).

1 This is a revised version of a paper that was originally presented at the Cowles Foundation Econometrics Conference, October 23–24, 1999, and subsequently presented at a conference in honor of the 80th birthday of Daniel McFadden at University of Southern California on July 28–29, 2017. We thank our discussants, Hal Varian (who discussed our paper at the McFadden conference) and the late Halbert White (who discussed our paper at a conference at Yale in 1999) as well as very helpful comments from Robert Engle and George Tauchen at the Yale conference. We also received helpful feedback from Michael Keane, Kenneth Wolpin, Frank Vella and other seminar attendees at subsequent presentations of this paper at the University of Pennsylvania, Boston University, Johns Hopkins, Rice, Penn State, the Universities of Chicago, Maryland, Pittsburgh, Virginia and Florida, the 2001 Conference of the Society for Computational Economics, and the 2001 Midwest Econometrics Group Annual Meeting. We are grateful for financial support from National Science Foundation, USA grant SES-9905145.

https://doi.org/10.1016/j.jeconom.2020.07.033
0304-4076/© 2020 Elsevier B.V. All rights reserved.

Please cite this article as: G. Hall and J. Rust, Estimation of endogenously sampled time series: The case of commodity price speculation in the steel market. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.07.033.

result from a probabilistic sampling rule that depends on the state of the process that generally includes other observed covariates x_t.

We focus on a particular example where p_t denotes the daily wholesale price of a standardized steel product. There are no formal markets or centralized exchanges where steel is traded. Instead nearly all steel transaction prices are a result of private bilateral negotiations between buyers and sellers, typically intermediated by traders or "middlemen" that are called *steel service centers* in this industry.[2] Even though there is no central record of daily transactions prices in the steel market, we do observe transaction prices for a particular firm — a steel service center that purchases large quantities of steel in the wholesale market for subsequent resale in the retail market. The endogenous sampling problem arises from the fact that the firm only records wholesale prices p_t on the days that it purchases steel. Knowledge of the wholesale price process is critical to the behavior of traders in the steel market, who attempt to "buy low and sell high" by strategically timing steel purchases at low prices for subsequent resale to retail customers at a markup.

Though we focus on endogenous sampling in the case of the steel market, the problem arises in many other contexts. Examples include financial applications where transaction prices are observed at randomly spaced intervals (see Aït-Sahalia and Mykland, 2001; Duffie and Glynn, 2004; Engle and Russell, 1998), and in marketing applications where the prices of goods that a household purchases are generally only recorded for the items the household purchased and on the dates it purchased them (see Allenby et al., 1996; Erdem and Keane, 1996; Erdem et al., 1999).

However we are not aware of econometric or statistical studies that are directly relevant for handling endogenous sampling problems in a time series context. The closest related work is on high frequency financial time series such as Potiron and Mykland (2017) who analyze diffusion processes where "an observation is generated whenever an auxiliary process called observation time process hits one of the two boundary processes, it is called the hitting boundary process with time process (HBT) model." They focus on the question of estimating the instantaneous covariance between two diffusion processes, using an approach pioneered by Hayaski and Kusuoka (2008) for estimating the covariance when the processes are sampled non-synchronously at arbitrary stopping times.[3] We focus on estimating parametric Markov processes observed in discrete time. In the steel example, under the assumption of optimality, the time between successive purchases (and thus observations of wholesale prices p_t) is also a stopping time: it is the first passage time to an "order region" defined by a generalization of the (S, s) inventory rule of Scarf (1959). We also observe retail prices p_t^r but these are sampled non-synchronously with p_t, and are observed only when sales occur.

Our work is also related to the literature on likelihood-based methods for correcting for endogenous sampling and missing data in cross-sectional and panel contexts (Heckman, 1981; Manski and McFadden, 1981; McFadden, 1997; Erdem et al., 1999). The latter paper uses a simulated likelihood approach to solve a missing data problem in a marketing application that is similar to the one in our paper: they have high frequency scanner panel data on consumer purchases but "only the prices of the items bought are recorded." (See p. 177). However these likelihood-based approaches for cross sectional and panel data are not directly applicable to the steel problem, where we observe purchases and sales by a single firm. Inference in this case requires econometric methods for time-series, and a likelihood-based approach to solving the endogenous sampling problem requires high dimensional numerical integration or simulation to "integrate out" the unobserved values of p_t on days where purchases do not occur. For example, for 3/4 inch plate steel, while we observe inventories over 1647 business days purchases occurred on only 237 days, so each evaluation of the likelihood function requires integration over 1410 dimensions.

We show how inference can be conducted using a consistent, less efficient, but computationally simpler method of simulated moments (MSM) estimator that avoids high dimensional numerical integrations required by likelihood-based approaches. The MSM estimator was introduced by McFadden (1989) to avoid the burden of numerical integration in maximum likelihood estimation of discrete choice models. He showed that it is possible to consistently estimate parameters of random utility models using simulations of unobserved utilities by finding parameters that best fit simulated choices to actual choices. McFadden's key insight was that simulation error averages out in large samples just as sampling error does, so MSM enables consistent estimation even with a single simulation draw of random utilities per observation. The general applicability of McFadden's idea to other contexts was quickly appreciated, and the asymptotic properties of MSM were shown to hold for time series as well as the cross-sectional case that McFadden first analyzed. In time series applications MSM is also known as simulated moments estimator (SME) (Lee and Ingram, 1991; Duffie and Singleton, 1993), but we are unaware of applications where it has been used to solve the endogenous sampling problem.

We use the MSM estimator to estimate the parameters of a truncated lognormal AR(1) model of the wholesale price processes for particular types of steel plate. These constitute the "parameters of interest" since knowledge of this process is a critical input to the firm's trading strategy. In particular, stationary AR(1) processes exhibit mean reversion that forms a basis for the intuitive "buy low, sell high" approach to trading. However the more successful the firm is as a

2 It is a puzzle why centralized exchanges exist for some commodities such as pork bellies, but not for steel. Rust and Hall (2003) develop a theory of intermediation in which the microstructure of trade in a commodity or asset is endogenously determined. Depending on the parameters of this model there are equilibria consistent with all trade occurring via a *market maker* on a centralized exchange, or all trade occurring via decentralized transactions with *middlemen*, or trade segmenting between middlemen and market makers. This theory can explain the variety of different trading institutions that we see in different markets, including the nonexistence of centralized exchanges for steel.

3 Hayaski and Kusuoka (2008) note that "Estimation problems of the diffusion parameter for diffusion-type processes based on discrete-time samples have been studied in statistics for a long time ... To the best of our knowledge, however, nonsynchronous cases have seldom been investigated."

ARTICLE IN PRESS

speculator (i.e. in strategically timing its purchases), the more serious are the potential biases that would result from failing to account for the endogeneity of the sampling process. In particular, the mean of observed purchase prices will be a downward biased estimate of the long run mean price of steel, and the estimated serial correlation in prices that are infrequently and endogenously sampled will be a downward biased estimate of the true daily serial correlation in prices.

MSM solves the endogenous sampling problem in a extremely straightforward manner: all it requires is the ability to simulate time series of steel purchases, sales, and inventories. These simulations are then sampled in exactly the same way as the observed data are sampled, similar to the way McFadden simulated choices from simulated random utilities. The MSM estimator is the parameter value that minimizes the distance between simulated moments and observed moments, where both the actual and simulated data are sampled according to the same sampling rule; namely, *wholesale prices are observed only on the days the firm purchases steel.*

Even though the moments relating to wholesale prices entering the MSM criterion (such as the mean price) are biased and inconsistent due to endogenous sampling, the fact that we can sample the data entering the simulated moments in the same way that the actual data are sampled can be used to establish its consistency. It should be apparent that although the details of how the time series data are sampled are case-specific and our analysis of endogenous sampling is specific this particular steel example, it should be easy how to apply MSM in other types of endogenous sampling problems that arise in a variety of other contexts. In each case it is necessary to formulate a model that is rich enough to simulate and sample data in the same way that the actual data are sampled.

The goal of our empirical work is not just on inference on the stochastic process for prices $\{p_t\}$ but also to test the assumption that the firm behaves as discounted expected profit maximizer. This paper originated from previous empirical and theoretical work (Hall and Rust 1999, 2003, 2007) on modeling the speculative trading and inventory investment decisions of a particular steel wholesaler. In this previous work we showed that the optimal speculative trading strategy for the firm (i.e. the strategy that maximizes its discounted profits) is a generalization of the classic (S, s) rule of Scarf (1959).

Let $\{p_t\}$ denote the stochastic process representing the *lowest price* offered by any seller of a particular steel product on day t. We assume this process is Markovian and that the firm observes p_t at each day t, but it only records p_t when it decides to place an order. Let q_t^o denote the quantity ordered on day t. The endogenous sampling rule can be stated as follows:

$$p_t \quad \text{is observed} \quad \Longleftrightarrow \quad q_t^o > 0.$$

It is notationally convenient to treat the endogenous sampling problem as a censored sampling problem: i.e., we set p_t to some arbitrary value such as $p_t = 0$ when $q_t^o = 0$, and let p_t equal the observed purchase price when $q_t^o > 0$. Note that we also observe the retail sales prices $\{p_t^r\}$ that the firm charges its customers and these prices are also endogenously sampled: the firm only records sales prices on days where sales occur (it does not record price quotes that customers reject). Since retail sales occur much more frequently than purchases on the wholesale market, retail price data $\{p_t^r\}$ can provide a key source of information for learning about $\{p_t\}$. However on the subset of days where both p_t and p_t^r are observed, we observe that markups $p_t^r - p_t$ are quite volatile, and vary by time, location, and type of the customer. In other words, there is considerable price discrimination in the retail market for steel. As a result the retail price of steel p_t^r is best regarded as a noisy and biased signal of the wholesale price p_t and therefore retail prices, while observed more frequently than wholesale prices, are not sufficient by itself for estimating the unknown parameters of the wholesale price process.[4]

MSM estimation under the hypothesis of optimality requires nested numerical solution of a dynamic programming (DP) problem that determines the firm's optimal trading strategy, and hence the endogenous sampling rule. The DP problem is re-solved for each trial value for the unknown parameter vector θ, and as a result, the MSM estimator can be computationally intensive. However significant computational savings can be achieved by exploiting special features of the solutions to these dynamic programming problems. Extending a seminal result by Scarf (1959) for a simpler class of inventory investment problems, Hall and Rust (2007) showed that the optimal speculative investment strategy for a fairly general class of commodity price speculation problems takes the form of a *generalized* (S, s) *rule.* In a generalized (S, s) rule, the thresholds S and s are functions of the current wholesale price p and a vector of other state variables x such as interest rates, demand shifters, and other variables that affect the firm's beliefs about future prices and sales levels. The functions $S(p, x)$ and $s(p, x)$ satisfy $S(p, x) \geq s(p, x)$. The lower band $s(p, x)$ is the firm's *order threshold:* it is optimal for the firm to place an order whenever its current inventory level q falls below $s(p, x)$. The upper band $S(p, x)$ is the firm's *target inventory level:* whenever the firm places an order to replenish its inventory, it orders an amount sufficient to insure that inventory on hand (the sum of the current inventory plus new orders) equals $S(p, x)$. The order threshold function $s(p, x)$

4 The assumption that $\{p_t\}$ is observed each day by the firm and evolves as an exogenous stochastic process (i.e. its realizations do not depend on actions of the firm) is one that we intend to relax in future work. Prices in the steel market are determined via bilateral negotiations: there is no central market place where the wholesale prices can be directly observed. Instead, in order to get price quotes, purchasing agents within the firm must communicate with steel producers or other intermediaries via telephone, fax, telex, or the internet. Thus the firm must search to find the best wholesale price. However this leads to additional endogeneity problems, since the best price the firm is able to negotiate depends on the intensity of its search/bargaining process, which could vary depending on the conditions it faces. We defer the difficult issues associated with modeling price search by the firm to future research. While this complicates the dynamic programming problem we present in Section 3, we believe that MSM estimation can still be applied.

Please cite this article as: G. Hall and J. Rust, Estimation of endogenously sampled time series: The case of commodity price speculation in the steel market. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.07.033.

encodes the endogenous sampling of $\{p_t\}$ since the firm only records the wholesale price p_t on the days where purchases occur. Therefore the sampling is described by the following threshold rule

$$p_t \quad \text{is observed} \quad \Longleftrightarrow \quad q_t < s(p_t, x_t). \tag{1}$$

Conditional on a purchase, we observe an order of size q_t^o given by

$$q_t^o = S(p_t, x_t) - q_t, \tag{2}$$

and $q_t^o = 0$ otherwise. Using the generalized (S, s) rule as our model of the endogenous determination of sampling dates, we propose estimators that are able to consistently estimate the unknown parameters of the $\{p_t\}$ process even though the process is endogenously sampled.

The hypothesis of optimality puts strong restrictions on the timing and purchase quantities of steel, and hence strong restrictions on how wholesale prices $\{p_t\}$ are endogenously sampled. If the assumption of optimality is incorrect, its imposition can result in distorted and inconsistent estimates of the AR(1) parameters for $\{p_t\}$ as well as other parameters of interest. Thus, we also formulate and estimate an *unrestricted* MSM estimator that does not impose the assumption of optimality. The unrestricted model replaces the sampling rule (1) with a probabilistic sampling rule: the firm purchases steel, and hence p_t is observed, with probability $P(q_t, p_t, x_t)$ where the quantity of steel purchased, q_t^o, is a realization from a distribution F, $q_t^o \sim F(q^o|q_t, p_t, x_t)$. If (P, F) are allowed to be arbitrary probability distributions, then it is easy to see that the unrestricted model nests the (S, s) model as a special case. We can view (P, F) as infinite-dimensional "nuisance parameters" to the extent that our main interest is to estimate the parameters of the AR(1) process $\{p_t\}$. However (P, F) are also of interest, since they capture the firm's strategy for buying steel. Borrowing from the literature on semi-parametric sieve estimation for weakly dependent data (Chen and Shen, 1998), we employ series estimators to approximate the unknown (P, F) using flexible parametric approximations using J parameters where J increases with the sample size. Specifically, we assume that P is given by a binary logit model, and F is a mixture of truncated lognormal distributions, with location and scale parameters given by a linear combination of J "basis functions" that depend on the state variables of the model, (p, q, x). When J and the number of mixture components M increase with the sample size at the right rate, MSM can consistently estimate the parameters of (P, F) in the presence of endogenous sampling. Thus, the unrestricted specification (which we also refer to as the *reduced-form* model), can be used to test the assumption that the firm's purchases of steel are governed by an (S, s) rule.

Section 2 describes our data, the steel company's problem, and the endogenous sampling problem that motivated this paper. Section 3 introduces a model of optimal commodity price speculation and inventory management and defines the generalized (S, s) trading strategy that forms the basis of our structural model of the endogenous sampling of wholesale steel prices. We also describe an unrestricted version of the model that relaxes the optimality assumption. Section 4 introduces the MSM estimators for the fully parametric (S, s) model and the semi-parametric model that relaxes the assumption of optimality. We summarize its asymptotic properties in a time-series setting when the data are endogenously sampled. Section 5 present the results of an empirical application of the MSM estimator to two plate steel products for which wholesale prices are assumed to evolve according to a univariate truncated lognormal AR(1) process. We estimate the unknown parameters of the price process and the unknown parameters affecting the firm's cost of purchasing and holding inventory. We then evaluate how well our generalized (S, s) trading strategy fits the data and use our results to infer the fraction of the firm's discounted profits are due to the markups it charges its retail customers, and the fraction that is due to pure commodity price speculation, i.e., its success in timing purchases in order to profit from "buying low and selling high." We reject the hypothesis that the firm is behaving optimally and show that the firm could significantly increase its profits by adopting an (S, s) trading strategy to optimize the timing and size of its steel purchases.

2. Steel data

An executive at a U.S. steel wholesaler provided us with the company's operations database that enables us to track the inventories and all sales and purchases of the 8900+ individual steel products that the company offers on a daily basis. The empirical results presented in Section 5 are based on data on every transaction the firm made between July 1, 1997 to December 31, 2003 (1647 business days) for two of its highest volume steel products, 3/4 and 1 inch plate steel. For each transaction we observe the quantity (number of units and/or weight in pounds) of steel bought or sold, the sales price, the shipping costs, and the identity of the buyer or seller.

Although this is an exceptionally clean and rich dataset, we only observe prices on the days the firm actually made transactions: the firm does not record any price information on days that it does not transact (either as a buyer or seller of steel). This shortcoming of our dataset is much more important for steel purchases than steel sales, since the firm purchases new steel inventory in the wholesale market much less frequently than it sells steel to its retail customers. Indeed, even for its highest volume products, it makes purchases only about once every two weeks. The (S, s) theory we present in Section 3 predicts that purchases are not made at random. Instead, the firm tends to make purchases when prices are low, so that the average price on the days the firm makes purchases will be lower than the average wholesale price on days the firm does not purchase. The exception to this general rule is that the firm may make purchases even when prices are relatively high if its inventories are low. Conversely, the firm may refrain from purchasing even if prices

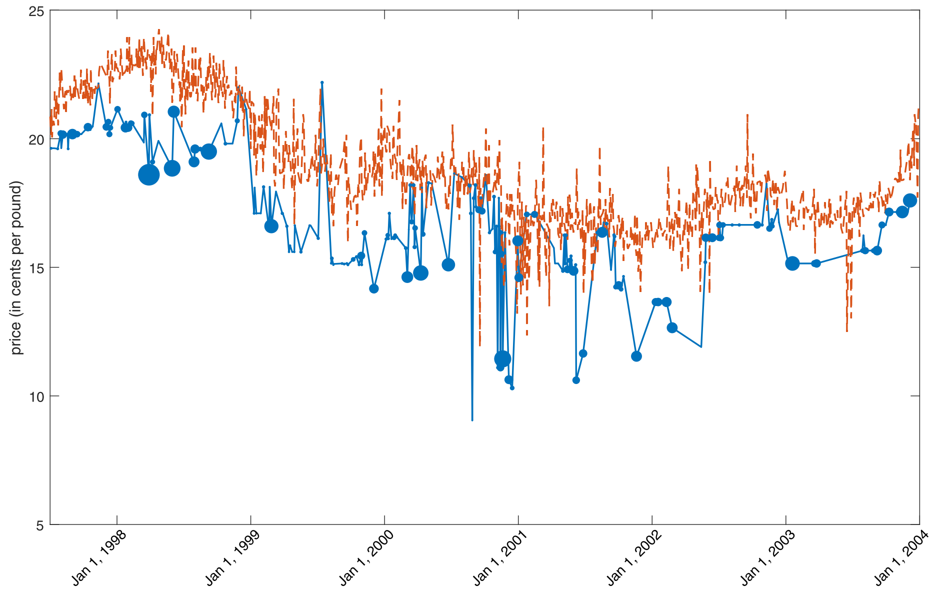


Fig. 1. Purchase prices (solid line) and retail prices (dashed line) of 3/4 inch plate. For the purchase price series, the size of the marker is proportional to the size of the purchase.

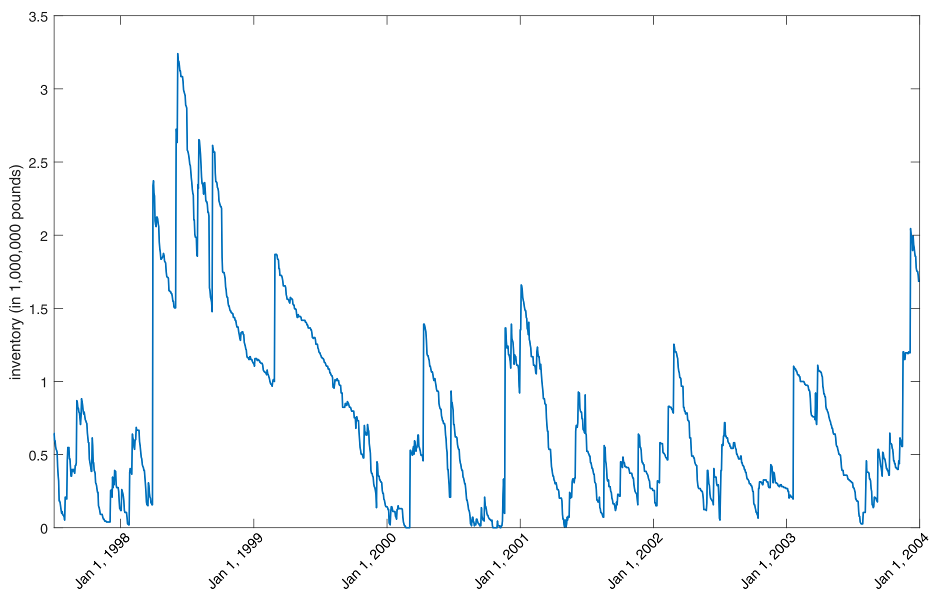


Fig. 2. Times series plot of inventory of 3/4 inch plate.

and inventories are low if it expects that the rate of retail sales will be depressed for a long period of time, say due to bad macroeconomic conditions. Thus, while the firm is attempting to "buy low and sell high," its purchase decisions involve a tradeoff among a number of different considerations.

We illustrate our data by plotting the time series of prices and inventories of 3/4 inch steel plate in Figs. 1 and 2. Steel plates are one of highest volume products sold by this firm. This type of plate is also a benchmark product within the industry since the prices of several other steel products are often computed as a function of its price. Unfortunately there are no exchanges for steel products: transactions in the steel market are done via private negotiations. The only price information comes from surveys of participants in the steel market that are reported as weekly or monthly averages of certain standardized products in trade publications such as *Purchasing Magazine* and *American Metal Market.* However the firm we study typically pays considerably different prices than those in the survey data.

As a result, in our plots of wholesale transaction prices in Fig. 1 (the lower curve with the large blue circles), we used straight line interpolations between observed purchase prices at successive purchase dates. The blue circle at each purchase date is proportional to the size of the firm's purchase in pounds. This gives us our first visual indication of the endogenous sampling problem. First, we see that even though we have 1647 observations on this firm, we observe purchases in the wholesale market on only 237 days. Second, the clustering of the blue dots at lower prices suggests that the firm is more likely to purchase large quantities of steel when wholesale prices are low, although other economic factors seem to be influencing the firm's purchase decisions as well. One key factor is the level of inventory: the firm tends to make large purchases when its inventory is low. We also see that even though wholesale prices continued to decline during 2000 and 2001, the firm's largest purchases of steel occurred during the "turning point" in prices in early 1998. The firm may have avoided making large purchases in late 2000 and 2001 due to economic uncertainties resulting from the "dot com crash" and the economic uncertainties following the 9/11/2001 attack on the U.S.

Overall, our interpolated plot of steel wholesale prices in Fig. 1 suggests that we should be wary of using the relatively small number of irregularly spaced observations to make inferences about the underlying law of motion for $\{p_t\}$. The observed purchase prices are unlikely to be representative of the unconditional mean level of prices in the wholesale market (especially if the firm is attempting to "buying low and sell high"), and the estimated serial correlation coefficient for these irregularly spaced transactions is unlikely to be a good estimate of the serial correlation coefficient between daily wholesale prices (assuming we were able to observe them).

Fig. 1 also plots the interpolated sequence of daily retail sales prices. Retail sales occur on about two out of every three business days, so the amount of interpolation in the retail price series is modest. The wholesale and retail prices move in a roughly parallel way, although there appears to be considerable day-to-day variation in retail prices. Retail prices are quoted net of transportation costs, but still much of the high frequency variation is due to observable factors. Athreya (2002) found that roughly 65% of the high frequency variation in retail prices can be explained by observable customer characteristics such as geographical location and past volume of purchases. The remaining 35% of the variation in retail prices appears to be due either to high frequency fluctuations in wholesale prices or to some sort of "informational price discrimination" in the retail market. Using the limited number of days on which both wholesale and retail prices are available, Chan (2001) found that at most 50% of the variation in retail prices can be explained by variations in the wholesale price of steel. This conclusion is possible due to the fact that on many days there are multiple retail sales to different customers. These findings suggest that a large share of the high frequency variation in retail prices can be ascribed to price discrimination, i.e. the firm charges higher prices to more impatient or poorly informed retail customers. We conclude that even though retail sales occur much more frequently than wholesale purchases, the fact that retail prices involve a number of other different considerations (including price discrimination based on observable and unobservable characteristics of the customer) suggest that the retail price is at best a noisy and (upward) biased signal of the underlying wholesale price.

Fig. 2 plots the evolution of inventories over the same period. Purchases of steel are easily recognizable as the discontinuous upward jumps in the inventory trajectories. As is evident from the saw-tooth pattern of the inventory holdings, the firm purchases the product much less frequently than it sells it. The firm's opportunistic purchasing behavior is evident for this product. As can be seen in Fig. 2 during the first ten months of the sample, from July, 1997 until March, 1998, the firm held relatively low levels of inventories at a time when the average price the firm paid for steel was about 20.5 cents per pound. However as the Asian financial crisis deepened, foreign steel producers began cutting their prices and aggressively increasing their exports. We see this clearly in our data, where in April 1998, wholesale prices dropped to 18.5 cents per pound. At that time the firm made a large purchase. As the price of steel continued to fall to historical lows during the remainder of 1998 the firm made a succession of large purchases that lead it to hold historically unprecedented high levels of inventories. We view this as clear evidence that the firm is attempting to profit from a "buy low, sell high" strategy.

3. Models of commodity price speculation

In this section introduces the two models of the firm's steel trading behavior that we will estimate and test empirically. Section 3.1 outlines a model of optimal commodity price speculation where the firm's trading strategy is governed by an (S, s) rule, and Section 3.2 discusses a "reduced form" model of trading that relaxes the assumption of optimality and allows firm behavior to be modeled using a much larger and flexible class of randomized purchasing strategies.

3.1. Optimal commodity price speculation: the (S, s) rule

Hall and Rust (1999, 2007) showed that in a broad class of commodity price speculation problems, the optimal trading rule is a generalized version of the classic (S, s) rule from inventory theory. Their work linked contributions by Arrow et al. (1951) and Scarf (1959), who first proved the optimality of (S, s) policies in inventory investment problems, to more recent work by Williams and Wright (1991), Deaton and Laroque (1992) and Miranda and Rui (1999) on the rational expectations commodity storage model. The fixed (S, s) thresholds derived by Scarf under the assumption of a constant price (cost) of procuring (producing) inventories is suboptimal in a speculative trading environment, where fluctuating prices affect both the optimal level of inventory S, and the threshold for purchasing new inventory s. Hall and Rust (1999,

2007) showed that the optimal trading strategy is generalized (S, s) rule where S and s are functions of state variables that include the wholesale price of steel p.[5]

Before we describe how the generalized (S, s) rule allows us to formulate and solve the problem of endogenous sampling of steel wholesale prices, we describe the notation and key assumptions underlying Hall and Rust's model of commodity price speculation. Then we formally define the (S, s) trading strategy, and show how in a broad class of models of speculation, the (S, s) rule constitutes the optimal strategy for "buying low and selling high." We assume that the firm can purchase unlimited quantities of steel at a time-varying wholesale price p_t that evolves according to a Markov transition density to be specified below. We assume that the firm subsequently sells this steel to retail customers at a retail price p_t^r that includes a randomly varying markup over the current wholesale price p_t (if we think of the firm as selling to different customers on different business days, this randomly varying markup is intended to be a "reduced-form" approach to capturing the pricing and price discrimination decisions by the firm).

On each business day t the following sequence of actions occurs:

1. At the start of day t the firm knows its inventory level q_t, the current wholesale price p_t, and the values of the other state variables x_t.
2. Given (q_t, p_t, x_t) the firm orders additional inventory q_t^o for immediate delivery.
3. Given (q_t, q_t^o, p_t, x_t) the firm sets a retail price p_t^r that is modeled as a random draw from a density $g(p_t^r|q_t+q_t^o, p_t, x_t)$.
4. Given $(q_t, q_t^o, p_t, p_t^r, x_t)$ the firm observes a realized retail demand for its steel, q_t^r, modeled as a draw from a distribution $H(q_t^r|p_t, p_t^r, x_t)$ with a point mass at $q_t^r = 0$.
5. The firm cannot sell more steel than it has on hand, so the actual quantity sold satisfies
$$q_t^s = \min\left[q_t + q_t^o, q_t^r\right]. \tag{3}$$
6. Sales on day t determine the level of inventories on hand at the beginning of business day $t + 1$ via the standard inventory identity:
$$q_{t+1} = q_t + q_t^o - q_t^s. \tag{4}$$
7. New values of (p_{t+1}, x_{t+1}) are drawn from a Markov transition density $f(p_{t+1}, x_{t+1}|p_t, x_t)$.

Note that we abstract from delivery lags and assume that the firm cannot backlog unfilled orders. Thus, whenever demand exceeds quantity on hand, the residual unfilled demand is lost. Thus, in addition to the censoring of the purchase and retail prices (p_t, p_t^r), we only observe a truncated measure of the firm's retail demand, i.e., we only observe the *minimum* of q_t^r and $q_t + q_t^o$ as given in Eq. (3). Since the quantity demanded has support on the $[0, \infty)$ interval, equation (3) implies that there is always a positive probability of a *stockout* given by:

$$\delta(q, p, p^r, x) = 1 - H(q|p^r, p, x). \tag{5}$$

Since retail sales occur much more frequently than purchases of new inventory, the retail sales price p_t^r provides an important source of information about the wholesale price p_t. Presumably for most transactions we should have $p_t^r \geq p_t$, reflecting nonnegative markups over the current wholesale price of steel.

The firm's expected sales revenue function, $ES(p, q, x)$ is the conditional expectation of realized sales revenue $p^r q^r$ given the current wholesale price p, quantity on hand q, and the observed information variables x. The firm's retail sales on date t is a random draw q_t^r from a conditional distribution $H(q_t^r|p_t^r, p_t, x_t)$ that depends on the retail price quote p_t^r, the current wholesale price p_t, and the values of the other observed state variables x_t. We assume that there is a positive probability $\eta(p^r, p, x) = H(0|p^r, p, x)$ that the firm will not make any retail sales on a particular day, so H can be represented by

$$H(q^r|p^r, p, x) = \eta(p^r, p, x) + [1 - \eta(p^r, p, x)] \int_0^{q^r} h(q|p^r, p, x)dq, \tag{6}$$

where h is a continuous strictly positive probability density function over the interval $[0, \infty)$. Given this stochastic "demand function," the firm's expected sales revenue $ES(p, q, x)$ is:

$$ES(p, q, x) = E\{\tilde{p}^r \tilde{q}^s|p, q, x\}$$

[5] This analysis extends previous results in the operations research literature such as Fabian et al. (1959), Kingman (1969), Kalymon (1971), Golabi (1985), Song and Zipkin (1993), Moinzadeh (1997), and Özekici and Parlar (1999) that prove the optimality of generalized versions of the (S, s) rule when the cost (price) of producing (procuring) new inventory fluctuates stochastically. While Hall and Rust (2007) are not the first to prove the optimality of generalized versions of the (S, s) rule, they extend the OR literature by making the connection between models of optimal inventory policies and models of storage and commodity prices. Moreover in the current paper we computationally solve and estimate our model. Thus we can formally compare the model's optimal policies to the inventory policies we see in the data. Besides the work noted above, the most closely related recent work that we are aware of is the ambitious paper by Aguirregabiria (1999) that models price and inventory decisions by a supermarket chain. A supermarket is similar to our steel wholesaler in that both types of firms hold inventories of a substantial number of different products, purchasing them in the wholesale market and selling their inventories at a markup to retail customers. The key difference is that prices in supermarkets are almost always posted so there is no direct price discrimination and there is presumably a larger "menu cost" to changing prices on a day by day basis. Aguirregabiria also did not directly address the endogenous sampling issue, using monthly price averages as proxies for underlying daily prices. For this reason we are unable to directly employ his innovative and ambitious approach to estimation of the steel problem.

$$= E\left\{\tilde{p}^r E\{\min[q, \tilde{q}^r]|p^r, p, q, x\}|p, q, x\right\} \\ = \int_0^\infty p^r[1-\eta(p^r, p, x)]\left[\int_0^q q^r h(q^r|p^r, p, x)dq^r + \delta(q, p^r, p, x)q\right] g(p^r|q, p, x)dp^r. \tag{7}$$

In order to state the per period profit function, we need to describe the costs that the firm incurs. The main cost is the cost of ordering new inventory, represented by the *order cost function* $c^o(q^o, p)$. We assume that the firm incurs a fixed cost $K \geq 0$ associated with placing new orders for inventory, which implies that $c^o(q^o, p)$ is given by

$$c^o(q^o, p) = \begin{cases} pq^o + K & \text{if} \quad q^o > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

The firm's remaining costs are summarized by the *holding cost function* $c^h(q, p, x)$. These costs include physical storage costs, and "goodwill costs" representing the present value of lost future business from customers whose orders cannot be filled due to a stockout. Goodwill costs can be viewed as the inverse of the "convenience yield" discussed in the commodity storage literature (Kaldor, 1939; Williams and Wright, 1991). In this case a convenience yield emerges from a desire to hold a buffer stock or precautionary level of inventories in order to minimize goodwill costs from stockouts. This allows the model to capture other reasons besides pure price speculation for holding inventories.[6] The firm's single-period profits π equals its sales revenues, less the cost of new orders for inventory $c^o(q^o, p)$ and inventory holding costs $c^h(q+q^o, p, x)$:

$$\pi(p, p^r, q^s, q+q^o, x) = p^r q^s - c^o(q^o, p) - c^h(q+q^o, p, x). \tag{9}$$

where $q^s = \min[q^r, q+q^o]$. Each period the firm chooses investment q^o_t given $\{p_t, q_t, x_t\}$ to maximize the discounted present value of profits:

$$V(p, q, x) = \max_{\{q^o\}} E\left\{\sum_{t=0}^\infty \beta^t \pi(p_t, p^r_t, q^r_t, q^o_t + q_t, x_t)\Big| p, q, x\right\}, \tag{10}$$

where $\beta = \exp\{-r/365\}$ is the firm's daily discount factor and r is the firm's annual discount rate and $\{q^o\}$ denotes a sequence of steel orders given by state-dependent decision rules that described in more detail below. The exponential formula for β accounts for the fact that our model has decisions being made at daily frequency. The value function $V(p, q, x)$ is given by the unique solution to Bellman's equation:

$$V(p, q, x) = \max_{0 \leq q^o \leq \overline{q}-q}\left[W(p, q+q^o, x) - c^o(q^o, p)\right], \tag{11}$$

where $\overline{q}$ is the firm's maximum storage capacity and

$$W(p, q, x) \equiv \left[ES(p, q, x) - c^h(q, p, x) + \beta EV(p, q, x)\right], \tag{12}$$

and EV denotes the conditional expectation of V given by:

$$EV(p, q, x) = E\{V(\tilde{p}, \max[0, q-\tilde{q}^r], \tilde{x})|p, q, x\}. \tag{13}$$

The optimal decision rule $q^o(p, q, x)$ is given by:

$$q^o(p, q, x) = \inf \mathop{argmax}_{0 \leq q^o \leq \overline{q}-q}\left[W(p, q+q^o, x) - c^o(q^o, p)\right]. \tag{14}$$

We invoke the inf operator in the definition of the optimal decision rule in Eq. (14) to handle the case where there are multiple maximizing values of q^o. We effectively break the tie in such cases by defining $q^o(p, q)$ as the *smallest* of the optimizing values of q^o.

In this model the variables q and q^o do not enter as separate arguments in the value function W given in (12): rather they enter as the sum $q+q^o$ as shown in Eq. (14). This symmetry property is a consequence of our timing assumptions: since new orders of steel arrive instantaneously, the firm's expected sales, inventory holding costs, and expected discounted profits only depend on the sum $q+q^o$, representing inventory on hand at the beginning of the period after new orders q^o have arrived. It follows that if the firm is holding less than its desired level of inventories $S(p_t, x_t)$ at the start of day t, it will only have to order the amount $q^o(p, q, x) = S(p, x) - q$ in order to achieve its target inventory level $S(p, x)$. Another way to see this is to note that when it is optimal for the firm to order, the optimal order level solves the first order condition:

$$\frac{\partial W}{\partial q^o}(p, q+q^o, x) = p. \tag{15}$$

[6] The firm obtains much of its steel from foreign sources. In the model orders occur instantaneously with certainty. In practice, however, delivery lags can be several months and the steel delivered can often be of lower quality than agreed on. The firm does have the option of refusing to take delivery if the steel is not of the quality promised. Having a buffer stock of inventories on hand reduces the cost to firm of exercising this option. Also foreign producers of steel do from time to time renege on previously negotiated deals, failing to deliver the amount of steel originally promised.

If W were strictly concave in q, there would be a unique value of $q+q^o$ that solves equation (15) for any value of p. Call this solution $S(p,x)$. It follows that $q+q^o=S(p,x)$, or $q^o(p,q,x)=S(p,x)-q$.

If $K>0$ the function $W(p,q,x)$ will not be strictly concave. However under fairly general conditions W is *K-concave* as a function of q for each fixed p.[7] Using the K-concavity property we can prove that whenever $q\geq s(p,x)$, it is not optimal to order: $q^o(p,q,x)=0$. When $q<s(p,x)$ the symmetry property implies that $q^o(p,q,x)=S(p,x)-q$ as discussed above. Hall and Rust (2007) proved:

Theorem 1. *Consider the function $W(p,q+q^o,x)$ defined in Eq. (12), where W is defined in terms of the unique solution V to Bellman's equation (11). Under appropriate regularity conditions given in Hall and Rust (2007), the optimal speculative trading strategy $q^o(p,q,x)$ takes the form of an (S,s) rule. That is, there exist a pair of functions (S,s) satisfying $S(p,x)\geq s(p,x)$ where $S(p,x)$ is the desired or target inventory level and $s(p,x)$ is the inventory order threshold, i.e.*

$$q^o(p,q,x)=\begin{cases}0 & \text{if}\quad q\geq s(p,x)\\ S(p,x)-q & \text{otherwise}\end{cases} \tag{16}$$

where $S(p,x)$ is given by:

$$S(p,x)=\underset{0\leq q^o\leq \overline{q}-q}{\operatorname{argmax}}\left[W(p,q^o,x)-c^o(q^o,p)\right] \tag{17}$$

and the lower inventory order limit, $s(p,x)$ is the value of q that makes the firm indifferent between ordering and not ordering more inventory:

$$s(p,x)=\inf_{q\geq 0}\{q|W(p,q,x)-pq\geq W(p,S(p,x),x)-pS(p,x)-K\}. \tag{18}$$

$s(p,x)$ and $S(p,x)$ are continuous functions of p for each x, and continuous in (p,x) if $W(p,q,x)$ is continuous in x.

3.2. Relaxing the optimality assumption

The assumption that firms behave optimally (i.e. they adopt strategies that maximize expected discounted profits) puts very strong restrictions on the timing and purchase quantities of steel, and hence strong restrictions on how wholesale prices $\{p_t\}$ are endogenously sampled. If the assumption of optimality is incorrect, its imposition can result in distorted and inconsistent estimates of the AR(1) parameters for $\{p_t\}$ as well as other parameters of interest. In this section we relax the assumption of optimality and describe a class of trading strategies that we refer to as the *unrestricted* or *reduced-form* model of steel trading, since we no longer require these strategies to maximize the expected discounted profits of the firm.

Consider a class of potentially suboptimal speculative trading strategies that include the optimal strategy as a special case. We will continue to impose a maintained assumption that the firm's behavior can be described by *randomized Markovian strategies* so its retail prices and purchase decisions can be expressed as probability distributions that depend on the current state (p_t,q_t,x_t). We assume the same timing as in the model of optimal speculation introduced in the previous section, and in particular we assume that at the start of each day t the firm observes the state (p_t,q_t,x_t) and chooses a quantity of steel q^o_t it wishes to purchase. We continue to assume that orders are delivered instantaneously, so after placing its order the firm has a total quantity of a particular steel product $q_t+q^o_t$ for sale to retail customers. We adopt the same "reduced form" specification for the retail price p^r_t, namely that retail prices are a draw from a conditional distribution $g(p^r_t|q^o_t+q_t,p_t,x_t)$. Given the realized value of p^r_t we also assume the same stochastic model of demand that we introduced above, namely the quantity sold q^s_t is given by $q^s_t=\min[q^o_t+q_t,q^r_t]$ where q^r_t is the retail demand on day t, which is a draw from the distribution $H(q^r_t|p_t,p^r_t,x_t)$ which has a mass point at $q^r_t=0$ to reflect the chance of zero demand on any particular day.

So to this point, the reduced form model is exactly the same as the model of optimal trading described in the previous section. The key difference between the reduced-form, unrestricted model of steel price speculation and the optimal model is to replace the deterministic optimal (S,s) rule for ordering steel given in Eq. (16) by a randomized rule defined by two conditional probabilities (P,F) where P represents the conditional probability that the firm places an order and F represents a conditional CDF for the size of the order placed.

Thus, we assume that the firm decides to place an order for steel with probability $P(q_t,p_t,x_t)$, and conditional on the decision to place an order, the quantity of steel ordered is a realization from a conditional probability distribution $F(q^o_t|q_t,p_t,x_t)$. If we allow (P,F) to be any conditional probabilities, then it is easy to see that the unrestricted specification of firm trading behavior includes the optimal (S,s) strategy as a special case by allowing P to be given by

$$P(q_t,p_t,x_t)=I\{q_t<s(p_t,x_t)\} \tag{19}$$

[7] A function $W(p,q):[\underline{p},\overline{p}]\times[0,\overline{q}]\rightarrow R$ is *K-concave* in its second argument q if and only if $-W(p,q)$ is *K-convex* in its second argument. More directly, $W(p,q)$ is *K-concave* in q iff $\exists K\geq 0$ such that for every $p\in[\underline{p},\overline{p}]$, and for all $z\geq 0$ and $b\geq 0$ such that $q+z\leq\overline{q}$ and $q-b\geq 0$ we have $W(p,q+z)-K\leq W(p,q)+z\,[W(p,q)-W(p,q-b)]\,/b$.

Please cite this article as: G. Hall and J. Rust, Estimation of endogenously sampled time series: The case of commodity price speculation in the steel market. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.07.033.

and F to be given by

$$F(q_t^o|q_t, p_t, x_t) = I\{q_t^o \geq S(p_t, x_t) - q_t\}. \tag{20}$$

In our empirical analysis we consider a flexible family of parametric distributions for (P, F) that can potentially include (19) and (20) as limiting special cases. For example, suppose P_σ is given by a binary logit,

$$P_\sigma(q_t, p_t, x_t) = \frac{1}{1 + \exp\{[q_t - s(p_t, x_t)]/\sigma\}} \tag{21}$$

where $\sigma \geq 0$ is a scale parameter. Then it is easy to see that $\lim_{\sigma \to 0} P_\sigma(q, p, x) = P(q, p, x)$ for almost all values of (q, p, x). Similarly, suppose $F_\sigma(q^o|q, p, x)$ is given by a lognormal distribution with location parameter $\mu(q, p, x) = S(p, x) - q$ and scale parameter σ. It is also easy to see that $F_\sigma \implies F$, in the sense of weak convergence of probability measures, as $\sigma \to 0$ for all (q, p, x).

Of course, we do not know the functional forms of the (S, s) bands $S(p, x)$ and $s(p, x)$ *a priori* without solving the firm's DP problem. However as shown above in Theorem 1, these are given by continuous functions of (p, x) under fairly weak assumptions on the firm's DP problem. Hence it is possible to approximate these functions by polynomial series. This motivates our use of flexible polynomial series specifications to approximate the logit probability of purchasing steel F given in Eq. (21) and for the (μ, σ) parameters of a lognormal specification for the quantity of steel purchased F.

We estimate a polynomial series specification for the probability of purchasing steel given by

$$P_{\gamma_P, J}(q, p, x) = \frac{1}{1 + \exp\left\{\sum_{j=1}^{J} \gamma_{P,j}\rho_j(q, p, x)\right\}} \tag{22}$$

where $\{\rho_j(q, p, x)\}$ are a family of "basis functions" such as tensor products of polynomials in (q, p, x) that can approximate arbitrary continuous functions, and thus arbitrary probabilities $P(q, p, x)$, including in the limit the degenerate probability given by the (S, s) rule, (19). Non-parametric density estimation of steel purchase quantities shows that F can be well approximated using mixtures of truncated normal distributions or truncated lognormal distributions with location and scale parameters $\{\mu(q, p, x), \sigma(q, p, x)\}$ that can each in turn approximated via polynomial series

$$\mu_{\gamma_\mu, J}(q, p, x) = \sum_{j=1}^{J} \gamma_{\mu,j}\rho_j(q, p, x) \tag{23}$$

$$\sigma_{\gamma_\sigma, J}(q, p, x) = \sum_{j=1}^{J} \gamma_{\sigma,j}\rho_j(q, p, x). \tag{24}$$

Let $\gamma_{F,J}$ denote the parameters of a mixture of lognormals where the location and scale parameters of the components lognormals in these mixtures are given by the series representations above and let $F_{\gamma_{F,J}}$ denote the resulting approximation to the true conditional CDF F. Other flexible families can be used, such as neural networks or splines. The key restriction is that these families are dense in the sense that if (P, F) is any pair of conditional probabilities, it can be approximated arbitrarily closely by a flexible parametric model $(P_{\gamma_{P,J}}, F_{\gamma_{F,J}})$ where the number of terms in the series approximations, J, and the number of mixture components M are sufficiently large. For any fixed J and number of mixture components M, we can easily simulate steel purchases from this approximate parametric model. This allows us to use the MSM estimator we introduce in the next section to estimate the parameters of interest, θ, which include the parameter for the AR(1) process for wholesale prices, the parameters of a parametric model of retail demand, and the firm's retail pricing rule, along with the parameters $(\gamma_{P,J}, \gamma_{F,J})$ characterizing our flexible "reduced form" model of firm purchasing behavior.

Thus, we have defined an expanding parametric family of randomized, Markovian purchasing strategies for the firm, indexed by J and M, that in the limit as $J \to \infty$ and $M \to \infty$, approximate arbitrarily closely any pair of conditional probabilities (P, F) including the limiting degenerate probability distributions characterizing the optimal (S, s) trading strategy. We now discuss how MSM can be used to solve the endogenous sampling problem, allowing us to estimate the probabilities (P, F) and the parameters of our model of commodity price speculation.

4. Method of simulated moments estimation

Both of the models of steel price speculation and inventory investment presented in Section 3 imply that $\{p_t, q_t, x_t\}$ evolves as a *controlled Markov process* where $\{p_t, x_t\}$ can be viewed as an "exogenous forcing process" whereas inventory $\{q_t\}$ is an endogenous variable whose law of motion is derived from the inventory accumulation identity (4). In addition to (p_t, x_t, q_t) we also observe (q_t^o, q_t^s, p_t^r) where q_t^o is the amount of steel ordered on day t, p_t^r is the retail price charged to customers on day t, and q_t^s is the amount sold to retail customers.[8] Thus, our interest is to model the evolution of the

[8] For simplicity, we assume that all sales on day t are done at a single posted price p_t^r. In actuality, the firm can sell to many different customers on a single day at individually negotiated prices. Thus the actual data we observe are more complex and a fully accurate treatment would necessitate modeling these individual negotiations and controlling for the individual customer characteristics and order sizes and delivery locations.

random vector $\{\xi_t\}$ given by

$$\xi_t \equiv (p_t, p_t^r, x_t, q_t^o, q_t^s, q_t) \tag{25}$$

and show that $\{\xi_t\}$ is a Markov process and provide sufficient conditions for it to have a unique invariant distribution. This invariant distribution plays a key role in the asymptotic properties of the MSM estimates of the model parameters θ as the number of observations $T \to \infty$.

An immediate issue that we need to confront is the potential *statistical degeneracy* of the (S, s) model: that is, unless there are unobservables included in the inventory speculation model, certain components of the vector ξ_t will be deterministic functions of the remaining components. For example, Theorem 1 implies that the optimal order quantity q_t^o is a deterministic function of (p_t, q_t, x_t) given in Eq. (16). Further, q_{t+1} is a deterministic function of ξ_t given in the inventory law of motion (4). The fact that q_t^o can be perfectly predicted from knowledge of (p_t, q_t, x_t) creates problems for statistical inference of the inventory model if we assume all components of x_t can be observed by the econometrician, since it is unlikely that any econometric model could perfectly predict every steel order made by the firm.

Inference is also complicated by frequently binding inequality constraints on inventory investment, q_t^o. This implies that it is not possible to use standard Euler equation methods to estimate the unknown parameters of the model via generalized method of moments (GMM) (see, e.g. Hansen, 1982). Note that Theorem 1 does yield a first order condition that could possibly provide a basis for a GMM estimator of the unknown parameters of the model:

$$\frac{\partial W}{\partial q}(p, S(p, x), x) - p = 0. \tag{26}$$

If we assume that there is additive measurement error ϵ in the wholesale price p, or assume that ϵ represents other unobserved (per unit) components of the cost of ordering new inventory, then it is tempting to treat Eq. (26) as an "Euler equation" and use GMM to estimate parameters of the model. However there are several big obstacles to this approach. First, we do not have a convenient analytical formula for the partial derivative of the value function, $\partial W/\partial q$. Second, even if the unconditional mean of ϵ is zero, the conditional mean of ϵ over those values of (p, ϵ) for which it is optimal to purchase (i.e. for which $q < s(p, x)$), is generally nonzero. Finally, there is the issue of endogenous sampling, and the fact that we observe purchases only on a relatively small subset of business days in our overall sample.

These problems motivate a search for an alternative approach that is capable of incorporating other information such as retail sales prices in order to improve our ability to make inferences about $\{p_t\}$. We start by describing how to apply the method of simulated moments (MSM) to estimate the "structural parameters" θ of an AR(1) model of wholesale prices as well as other parameters of the steel inventory speculation model under the assumption of optimality, where behavior is governed by the (S, s) rule described in Section 3.1. Then we discuss additional issues involved in non-parametric estimation of the "nuisance parameters" (P, F) as well as the "parameters of interest" θ_I (a subset of the structural parameter vector θ) that define the unrestricted specification that relaxes the assumption of optimality.

4.1. Asymptotics of MSM under the optimality assumption

In the context of our steel example, the favorable asymptotic properties of McFadden's MSM estimator carry over to a time-series setting, except that the Law of Large Numbers and Central Limit Theorem for serially correlated data need to be used, rather than ones for *IID* data that were appropriate in McFadden's analysis. Implementation of the MSM estimator is straightforward. First we calculate sample moments using the censored observations in the data, i.e. with $p_t = 0$ when $q_t^o = 0$. Then we generate one or more simulated realizations of the (S, s) model for a given trial value θ of the unknown parameter vector. We define $\hat{\theta}_T$ as the value of θ that minimizes a quadratic form in the difference between the sample moments for the actual data and the sample moments of the simulated data, where the simulated data has been censored in exactly the same fashion as the actual data, i.e. we set $p_t = 0$ whenever the simulated value of $q_t^o = 0$. Thus even though various moments based on censored data may be biased and inconsistent estimators of the corresponding moments of the ergodic process in the absence of censoring, this does not prevent us from deriving a consistent MSM estimator of θ^*.

The MSM estimator is easiest to implement computationally in cases where it is possible to construct *smooth simulators* i.e. where the simulated observations can be recursively constructed to be continuously differentiable functions of the structural parameters θ of the model. However while it is easy to construct a smooth simulator for the wholesale price process $\{p_t\}$ itself, the MSM estimator requires us to simulate a vector of endogenous variables, including orders of steel q_t^o and inventories q_t, and the nature of the (S, s) policy generates inherent discontinuities in the quantity variables since it implies that $q_t^o = 0$ when $s(p_t, x_t, \theta) \le q_t$ but q_t^o discontinuously jumps to $q_t^o = S(p_t, x_t, \theta) - q_t$ when $s(p_t, x_t, \theta) > q_t$. However the discontinuity along the lower $s(p, x)$ band is of measure zero within the full state space, so the Continuous Mapping Theorem implies that for any continuous, bounded mapping $h : R^L \to R^N$ we have $E\{h(\tilde{\xi})|\theta\} = \int_{\xi'} h(\xi')\psi(\xi'|\theta)d\xi'$ will be a continuous function of θ provided the transition probability λ is continuous in θ. Under Assumption 2, $E\{h(\tilde{\xi})|\theta\}$ will also be continuously differentiable in θ, which is needed to establish the asymptotic normality of the MSM estimator, whereas only continuity is required (along with additional assumptions guaranteeing geometric ergodicity of the $\{\xi_t(\theta)\}$ process uniformly for $\theta \in \Theta$ given in Duffie and Singleton, 1993) to establish the consistency of the MSM estimator.

Thus, the asymptotics of the MSM estimator are robust to discontinuities in simulations of the model with respect to the parameters θ for the same reason that McFadden's original MSM estimator is also robust to potential discontinuities

Please cite this article as: G. Hall and J. Rust, Estimation of endogenously sampled time series: The case of commodity price speculation in the steel market. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.07.033.

arising from small changes in the parameter values that cause simulated discrete choices to discontinuously jump from one value to another. The asymptotic properties depend on the true choice probability, which is smooth in the parameters, and this is also the case in our application.

The main complexity caused by discontinuities is not in the asymptotics (which are still $\sqrt{T}$ consistent and asymptotically normal in the presence of discontinuities in the simulated realizations of $\{p_t, x_t\}$) but rather in terms of computation of the MSM estimator, since the discontinuities in the simulations can lead to an estimation criterion that is locally flat and thus difficult to optimize. However by employing the ideas of Frazier and Zhu (2017) including the use of automatic differentiation, the practical difficulties caused by discontinuities in the MSM objective function can be significantly ameliorated.

The asymptotic properties of the MSM estimator depend on a key assumption of *ergodicity* of the inventory investment model, namely that there is a unique invariant distribution $\psi(\xi|\theta)$ for each $\theta \in \Theta$.

Assumption 1. For any $\theta \in \Theta$ the corresponding controlled Markov process $\{\xi_t(\theta)\}$ from the solution to the inventory investment problem (11) is ergodic with unique invariant distribution $\psi(\xi|\theta)$ given by:

$$\psi(\xi'|\theta) = \int \Lambda(\xi'|\xi,\theta)\psi(\xi|\theta)d\xi, \tag{27}$$

where Λ is the CDF corresponding to the transition density λ.

Assumption 1 can be established via lower level assumptions on the primitives of the model following the uniqueness and existence criterion in Futia (1982). For example it follows in the case where there are no x_t variables in the model and $\{p_t\}$ evolves according to a univariate AR(1) truncated lognormal process given in Eq. (45) of Section 5.1, when the autoregressive coefficient λ_p is in the unit interval $[0, 1)$. When there are x_t variables in the model, then due to our assumption that $\{p_t, x_t\}$ is jointly Markovian, Assumption 1 will follow from appropriate ergodicity assumptions on this joint process.

However we need a stronger result than just existence of a unique invariant distribution $\psi(\xi|\theta)$ for each $\theta \in \Theta$. We also require this invariant distribution to be differentiable in its parameters and also that expectation of any bounded, continuous function $h : R^L \to R^N$ (where L is the dimension of the Markov process, $\{\xi_t\}$, and N is the number of moments) we have

$$\frac{\partial}{\partial\theta}E\{h|\theta\} = \frac{\partial}{\partial\theta}\int h(\xi')\psi(d\xi'|\theta) \tag{28}$$

exists and is a continuous function of $\theta \in \Theta$. Vázquez-Abad and Kushner (1992) provide sufficient conditions for differentiability to hold and the key condition, in addition to uniqueness of the invariant measure, is that the collection of invariant distributions $\{\psi(\xi'|\theta)|\theta \in \Theta\}$ is tight. Tightness follows automatically when the state space is compact, which we assume. It is easy to see that differentiability property of the ergodic distribution in Eq. (28) holds when $\{p_t\}$ is a stationary AR(1) process. Due to space constraints, we do not specify the lower level assumptions necessary to establish the differentiability condition for the full process $\{\xi_t\}$(28). Instead we simply make

Assumption 2. For any $\theta \in \mathrm{int}(\Theta)$ and any bounded, continuous function $h : R^L \to R^N$,

$$\begin{aligned} E\{h|\theta\} &= \int_{\xi'} h(\xi')\psi(d\xi'|\theta) \\ E\{h(\tilde{\xi})|\xi,\theta\} &= \int_{\xi'} h(\xi')\lambda(\xi'|\xi,\theta)d\xi' \end{aligned} \tag{29}$$

are continuously differentiable in θ.

Assumptions 1 and 2, along with several additional assumptions to be introduced in the next section, are the key to establishing the asymptotic properties of the MSM as $T \to \infty$ where T is the number of time periods we observe the single steel firm that is the basis for or empirical analysis. In the next section we show how the MSM estimator is able to conveniently and easily handle the complexities of the endogenous sampling of the wholesale price process $\{p_t\}$.

The MSM estimator is based on finding the $D \times 1$ vector of parameter values, $\hat{\theta}_T$, that best fits a $M \times 1$ vector of moments of the observed process:

$$h_T \equiv \frac{1}{T}\sum_{t=1}^{T} h(\xi_t, \xi_{t-1}), \tag{30}$$

where $M \geq D$ and h is a known function of (ξ_t, ξ_{t-1}) that determines the moments we wish to match. We include ξ_t and its lag ξ_{t-1} as arguments of h in order to handle situations where we are trying to fit moments such as means and covariances of the components of ξ_t. It is straightforward to allow moments that involve more than one lag: we only include a single lagged value of ξ_t in our presentation below for notational simplicity.

By Assumption 1, the process $\{\xi_t\}$ is ergodic so that, with probability 1, h_T converges to a limit $E\{h(\xi', \xi)\}$ where the expectation is taken with respect to the ergodic distribution of (ξ', ξ) (i.e. the limiting distribution of (ξ_{t+1}, ξ_t) as $t \to \infty$). Under suitable additional regularity conditions, a central limit theorem will hold for h_T, i.e. we have

$$\sqrt{T}\,[h_T - E\{h\}] \Longrightarrow N(0, \Omega(h)), \tag{31}$$

where

$$\Omega(h) = \sum_{j=-\infty}^{\infty} E\left\{(h(\xi_{t+1}, \xi_t) - E\{h\})(h(\xi_{t+j+1}, \xi_{t+j}) - E\{h\})'\right\}, \tag{32}$$

where the expectations in (32) are taken with respect to the ergodic distribution of (ξ', ξ). Formula (32) is the standard formula for the variance of a sum for a serially correlated process when the number of elements in the sum tend to infinity, and is the same formula given in Assumption 3 of Duffie and Singleton (1993). To operationalize the infinite sum of covariances at all leads and lags in our empirical work, we employ the Newey and West (1987) estimator of $\Omega(h)$ in our empirical work, since as Duffie and Singleton note, the assumptions that insure the geometric ergodicity of the process $\{\xi_t\}$ also imply the α-mixing property necessary to insure the consistency of the Newey–West estimator.

Now assume it is possible to generate simulated realizations of the $\{\xi_t\}$ process for any candidate value of θ, and that this process is sampled/censored in exactly the same way as the observed $\{\xi_t\}$ process is sampled/censored, i.e., with $p_t = 0$ when $q_t^o = 0$. These simulations depend on a $T \times 1$ vector, u, of *IID* $U(0, 1)$ random variables that are drawn once at the start of the estimation process and held fixed thereafter in order for the estimator to satisfy a stochastic equicontinuity condition necessary to establish the consistency and asymptotic normality of the MSM estimator. We will consider simulated processes of the form

$$\{\xi_t(\{u_s\}_{s\le t}, \theta, \xi_0)\}, \quad t = 2, \ldots, T \tag{33}$$

where for each $t > 1$, $\xi_t(\{u_s\}_{s\le t}, \theta, \xi_0)$ is recursively defined to depend on previous simulated values for periods $s \le t$. The notation $\{u_s\}_{s\le t}$ reflects the fact that the simulated process is *adapted* to the realization of the $\{u_t\}$ process, i.e. the first t realized values of $\{\xi_t(\{u_s\}_{s\le t}, \theta)\}$ depend only on the first t realized values of $\{u_s\}$ and not on subsequent realized values of u_s for $s > t$. Note that we allow the simulated process to depend on the first value ξ_0 of the observed data as an initial condition.

Now consider using a single simulated realization of $\{\xi_t(\{u_s\}_{s\le t}, \theta, \xi_0)\}$ to form a simulated sample moment $h_T(\{u_s\}_{s\le T}, \xi_0, \theta)$ given by

$$h_T(\{u_s\}_{s\le T}, \xi_0, \theta) = \frac{1}{T}\sum_{t=1}^{T} h\left(\xi_t(\{u_s\}_{s\le t}, \theta, \xi_0), \xi_{t-1}(\{u_s\}_{s\le t-1}, \theta, \xi_0)\right). \tag{34}$$

Let $(\{u_s^1\}_{s\le T}, \ldots, \{u_s^n\}_{s\le T})$ denote n *IID* $T \times 1$ sequences of $U(0, 1)$ random vectors used to generate the n independent realizations of the endogenously sampled process $\{\xi_t(\{u_s^i\}_{s\le t}, \theta, \xi_0)\}$, $i = 1, \ldots, n$. Define $h_{n,T}(\theta)$ as the average of r independent time averages $h_T(\{u_s^i\}_{s\le T}, \xi_0, \theta)$

$$h_{n,T}(\theta) = \frac{1}{n}\sum_{i=1}^{n} h_T(\{u_s^i\}_{s\le T}, \xi_0, \theta). \tag{35}$$

Definition. The **method of simulated moments (MSM) estimator** $\hat{\theta}_T$ is given by:

$$\hat{\theta}_T = \underset{\theta\in\Theta}{argmin}(h_{n,T}(\theta) - h_T)' W_T (h_{n,T}(\theta) - h_T), \tag{36}$$

where W_T is an $M \times M$ positive definite weighting matrix.

In order to simplify the asymptotic analysis, we initially assume that we have a correct parametric specification of the endogenous sampling problem. That is we make

Assumption 3. The parametric model introduced in Section 3 is correctly specified, i.e., there is a $\theta^* \in \Theta$ such that:

$$\{\xi_t(\{u_s\}_{s\le t}, \theta^*, \xi_0)\} \sim \{\xi_t|\xi_0\} \tag{37}$$

that is, when $\theta = \theta^*$, the simulated sequence initialized from the observed value ξ_0 has the same conditional probability distribution as the observed sequence $\{\xi_t\}$ given ξ_0.

We now sketch the derivation of the asymptotic distribution of the MSM estimator, listing the key assumptions and showing how its asymptotic variance depends on the number of simulations n. By Assumption 2, we have that for any $\theta \in \Theta$ the simulated process $\{\xi_t(\{u_s\}_{s\le t}, \theta, \xi_0)\}$ is ergodic with unique invariant probability $\psi(\xi|\theta)$. Define the function $E\{h|\theta\}$ by

$$E\{h|\theta\} = \int h(\xi', \xi)\lambda(\xi'|\xi, \theta)\psi(d\xi|\theta)d\xi'. \tag{38}$$

By Assumption 2, its gradient $\nabla E\{h|\theta\}$ exists and is continuous in θ for $\theta \in \text{int}(\Theta)$.

Assumption 4. θ^* is **identified**; that is, if $\theta \neq \theta^*$, then $E\{h|\theta\} \neq E\{h|\theta^*\} = E\{h\}$. Furthermore, $\text{rank}(\nabla E\{h|\theta\}) = D \leq N$ and $\lim_{T\to\infty} W_T = W$ with probability 1 where W is a $M \times M$ positive definite matrix.

The consistency of the MSM estimator can be established by providing appropriate regularity conditions under which the simulated process is uniformly ergodic, i.e.

$$\lim_{T\to\infty} \sup_{\theta\in\Theta} \left|(h_{n,T}(\theta) - h_T)' W_T (h_{n,T}(\theta) - h_T) - (E\{h|\theta\} - E\{h|\theta^*\})' W (E\{h|\theta\} - E\{h|\theta^*\})\right| = 0. \tag{39}$$

Assumption 4 guarantees that the unique minimizer of $(E\{h|\theta\} - E\{h|\theta^*\})' W (E\{h|\theta\} - E\{h|\theta^*\})$ is θ^*, and this combined with the uniform consistency result implies the consistency of $\hat{\theta}_T$. The asymptotic normality of $\hat{\theta}_T$ can be established by the arguments given in Duffie and Singleton (1993). Similar to the literature on GMM estimation, the optimal weight matrix is $W = [\Omega(h, \theta^*)]^{-1}$ and it results in an MSM estimator with minimal variance, and in that case the asymptotic distribution of $\hat{\theta}_T$ is given by

Theorem 2. *Consider the MSM estimator $\hat{\theta}_T$ formed using a weighting matrix W_T equal to the inverse of any consistent estimator of $\Omega(h, \theta^*) = \Omega(h)$ such as the Newey–West estimator. Then under Assumptions 1 to 4, we have:*

$$\sqrt{T}[\hat{\theta}_T - \theta^*] \Longrightarrow N(0, (1 + 1/n)\Lambda^{-1}) \tag{40}$$

where:

$$\Lambda = \left[\nabla E\{h|\theta^*\}' [\Omega(h, \theta^*)]^{-1} \nabla E\{h|\theta^*\}\right]. \tag{41}$$

The most important point to note about this result is that it has the same form as the asymptotic variance that McFadden (1989) derived in the *IID* sampling case, namely the penalty to forming an MSM estimator using only a single realization $n = 1$ of the endogenously sampled process $\{\xi_t(\{u_s\}_{s\leq t}, \theta, \xi_0)\}$ is fairly small. The variance of the resulting estimator is only twice as large as an estimator that computes the expectation of $h_T(\{u\}, \theta)$ exactly, such as would be done via Monte Carlo integration when $n \to \infty$.

The MSM estimator can be implemented in practice by solving

$$\hat{\theta}_T = \underset{\theta\in\Theta}{argmin} (h_{S,T}(\theta) - h_T)' \left[\hat{\Omega}(h)\right]^{-1} (h_{S,T}(\theta) - h_T), \tag{42}$$

where $\hat{\Omega}(h)$ is the Newey and West (1987) estimator of $\Omega(h)$ given in Eq. (32). Thus, it is not necessary to recompute the optimal weighting matrix $\Omega(h, \theta)$ each time the parameter θ is updated, but instead we can pre-compute the empirical covariance matrix of the moments $\hat{\Omega}(h)$ using the Newey–West estimator at the start of the estimation process and the inverse of this matrix is the required consistent estimator of the optimal weighting matrix that can be used for the entire estimation process. Further details on how we set the bandwidth (i.e. number of lags) to compute $\hat{\Omega}(h)$ are provided in Section 5.3.

4.2. Asymptotics of MSM when the optimality assumption is relaxed

The (S, s) model that we estimate in Section 5.1 is fully parametric and depends on a $D \times 1$ vector of unknown structural parameters θ to be estimated. When we relax the optimality assumption, the model becomes semi-parametric: the parameters of the unrestricted or "reduced-form" model of firm behavior introduced in Section 3.2 consists of a $D_1 \times 1$ parameter vector θ_I together with the unknown functions (P, F), where P is the conditional probability that the firm chooses to purchase steel and F is the conditional probability distribution for the amount of steel purchased. The parameter vector θ_I can be regarded as "the parameters of interest" and it is $D_1 \times 1$ a subvector of the $D \times 1$ structural parameter vector θ containing the parameters of the AR(1) process for wholesale prices $\{p_t\}$, as well as the unknown parameters of the density governing the firm's retail pricing strategy $g(p_t^r|q_t + q_t^o, p_t, x_t)$ and the parameters of the distribution for retail demand, $H(q_t^r|p_t, p_t^r, x_t)$. Other components of the structural parameter vector θ such as the firm's discount factor, and the costs of holding inventory and placing orders are not identified in the unrestricted model and thus not included in the subvector θ_I.

It is generally not possible to estimate (P, F) directly as infinite-dimensional nuisance parameters along with the parameters of interest θ_I by MSM. However it is possible to consistently estimate them by MSM using the method of *sieves* (Grenander, 1981). A sieve is an expanding sequence of finite-dimensional parameterized subsets of the overall infinite-dimensional parameter space that has the property of denseness, i.e. any element in the infinite-dimensional space can be expressed as a limit of a sequence of approximations that are restricted to be elements of these finite-dimensional parametric subsets.

In the example we study here, the sieve consists of the finite-dimensional subset of parametric approximations $(P_{\gamma_P,J}, F_{\gamma_F,J})$ to (P, F) that depend on series approximations (22) to P and a series approximation to F involving a mixture of M truncated lognormal distributions with (μ, σ) location and scale parameters that are expressed as a series of an appropriate set of known "basis functions" $\{\rho_j(p, q, x)\}$ given in Eqs. (23) and (24). In Section 5.2 we use a set of basis functions equal to tensor products of ordinary polynomials in p and q (e.g. $1, p, p^2, \ldots$ and $1, q, q^2, \ldots$) since we have no x variables in that specification.

Ai and Chen (2003) and Newey and Powell (2003) derived asymptotic properties of "sieve minimum distance" (SMD) estimators of that include MSM estimators as a special case. They show that it is possible to estimate the finite dimensional parameters of interest θ_I, at standard rates (i.e. $\sqrt{T}$ where T is the sample size), via the method of sieves even in the presence of infinite dimensional nuisance parameters (P, F). The sieve estimator of these latter parameters is consistent but converges at rates slower than $\sqrt{T}$. Though these results apply to *IID* data, Chen and Shen (1998) have established similar results for time series data, when the observations are weakly dependent (i.e. are strictly stationary and satisfy a β-mixing condition).

To paraphrase these general results to our application, suppose the number of mixture components M_T and series terms J_T increases with the sample size at the right rate with the sample size T. Then the MSM estimator of the parameter vector $(\hat{\theta}_{I,T}, \hat{\gamma}_{P,J_T}, \hat{\gamma}_{F,J_T}, \hat{\omega}_{M_T})$ (where $\hat{\omega}_{M_T}$ is an estimator of the M_T mixture weights for the mixture of lognormal representation for F) will result in a consistent estimator of (θ_I, P, F) and $\sqrt{T}[\hat{\theta}_{I,T} - \theta_I]$ will converge in distribution to a zero mean normal random variable even though the SMD estimator $(P_{\hat{\gamma}_{P,J_T}}, F_{\hat{\gamma}_{F,J_T}})$ converges to (P, F) at slower than $\sqrt{T}$.

The conditions on the rates at which J_T and M_T must increase with T and the rate of convergence of the sieve estimator of (P, F) are quite technical and depend on a host of lower level assumptions, including a calculation based on the "size" of the infinite-dimensional space of (P, F) functions as measured by quantity known as the metric entropy with bracketing (see Chen, 2007). It is quite difficult to verify these conditions and provide the necessary lower level assumptions for these general results to apply in our application. This is a task we leave to future work. In our empirical work, we determined the number of terms J_T in the series approximations and the number of terms M_T in the mixture model approximation to F empirically, by increasing these values until further terms are insignificant and fail to increase the MSM objective function. Though it has not been established in the SMD case, Koenker (1988) showed that data-driven procedures for determining the number of terms J_T to enter in a series estimator similar to the one we employ in Section 5 do result in the number of parameters growing at the right rate with the sample size (i.e. as predicted by asymptotic theory for series estimation of regression models). Alternatively, we can interpret the unrestricted model as a fully parametric model. Thus, apart from the usual concerns about "specification searching" and pre-testing and model selection biases, it seems reasonable to assume that the MSM estimator of the parameters of interest in the unrestricted model specification will be well approximated by an asymptotic normal distribution. But to make things clear, we state this as

Assumption 5. Let $\hat{\theta}_{I,T}$ be the MSM estimator of the parameters of interest of the unrestricted model where the infinite-dimensional nuisance parameter (P, F) is simultaneously estimated by MSM using a sieve with (J_T, M_T) terms. The values of J_T and M_T are determined from a data-driven sequential estimation and testing procedure J_T is sequentially increased by adding higher order polynomial series terms, and M_T is increased by adding additional mixture model components until the newly added terms and mixture components are no longer significant according to conventional t-tests of significance of the coefficients and mixture weights as in Koenker (1988). Then we have

$$\sqrt{T}[\hat{\theta}_{I,T} - \theta_I] \Longrightarrow N(0, V) \tag{43}$$

where V is the $D_1 \times D_1$ submatrix of the asymptotic covariance matrix (41) corresponding to the identified parametric components of the unrestricted model.

5. Empirical application

To illustrate the MSM estimator, we consider a special case of the profit maximization model, introduced in Section 3.1, in which there are no additional state variables, x. In this case, the (S, s) bands are only functions of the current wholesale price, $S(p_t)$ and $s(p_t)$. We estimate the profit maximization model twice using actual data for two products, both large construction grade plates, from the steel service center.

We then re-estimate the two models relaxing the optimality assumption as discussed in Section 3.2. Since we are considering the case in which there are no additional state variables, the firm places an order for steel with probability $P(q_t, p_t)$, and conditional on the decision to place an order, the quantity of steel ordered is a realization from a conditional probability distribution $F(q_t^o|q_t, p_t)$. Using the estimated profit maximization models and their non-optimizing analogues, we decompose the firm's profits by product into four components. We use this decomposition to infer the share of the firm's profits that are due to markups paid by retail customers and the share due to price speculation. We also use this decomposition to compare the general manager's purchasing decisions to the model's optimal (S, s) rule and potentially suboptimal unrestricted trading rules (P, F).

5.1. A special case of the profit maximization model

Consider a version of the model in which the firm's general manager solves the following profit maximization problem:

$$\max_{\{q_t^o\}} E \sum_{t=0}^{\infty} \beta^t \left\{ p_t^r q_t^s - c^o(q_t^o, p_t) - c^h(q_t + q_t^o, p_t) \right\} \tag{44}$$

subject to (3), (4), and (8), and where $c^h(q_t+q_t^o, p_t) = \phi_1(q_t+q_t^o)+\phi_2(q_t+q_t^o)^2$. As before, the manager takes the wholesale price p_t and quantity demanded q_t^r as given. The manager knows p_t before deciding q_t^o. The manager then draws q_t^r. The holding cost function is quadratic so the marginal holding cost is increasing in the level of inventories.

We assume the wholesale price evolves according to a truncated lognormal $AR(1)$ process:

$$log(p_{t+1}) = \mu_p + \lambda_p log(p_t) + w_t^p \tag{45}$$

where w_t^p is an IID $N(0, \sigma_p^2)$ sequence. The firm sets the retail price using a fixed markup rule over the current wholesale price:

$$p_t^r = \frac{1}{\exp(\alpha_0^c + \alpha_0^s p_t)} + \frac{1}{\exp(\alpha_1^c + \alpha_1^s p_t)}\, p_t. \tag{46}$$

The firm draws a quantity demanded q_t^r each period from a mixed truncated lognormal distribution conditional on p_t. That is, with probability η, $q_t^d = 0$, and with probability $1-\eta$, q_t^d is drawn from a truncated normal distribution with location parameter $\mu_q(p) = \mu_p - \varsigma log(p_t)$. Let η be a function of the current price,

$$\eta = \frac{1}{\exp(\eta_c + \eta_s p_t)}. \tag{47}$$

Let θ^{pm} denote the $(D \times 1)$ parameter vector, $\{r, \eta_c, \eta_s, \alpha_0^c, \alpha_0^s, \alpha_1^c, \alpha_1^s, \mu_q, \sigma_q, \varsigma, K, \phi_1, \phi_2, \lambda_p, \mu_p, \sigma_p\}$.

5.2. A "reduced form" version of the model

We compare our profit maximization model to an alternative model in which we replace the optimal decision rule for q_t^o with a rule defined by two conditional probabilities, (P, F). In each period the probability the firm makes a purchase is given by the logistic function,

$$P(q_t, p_t) = Pr(q_t > 0|q_t, p_t) = \frac{1.0}{1.0 + \exp(\gamma_0 + \gamma_1 q_t + \gamma_2 p_t + \gamma_3 p_t^2 + \gamma_4 I(q_t < 30))}. \tag{48}$$

where $I(q_t < 30)$ is an indicator function which takes on a value of 1 if current inventories are below 30,000 pounds and 0 otherwise. Conditional on the firm making a purchase, the size of the purchase is a realization from the conditional probability distribution, $F(q_t^o|q_t, p_t)$. We approximate F with a lognormal distribution with a location parameter, $\mu_{q^o}(q, p)$, that depends on the firm's current level of inventories, q_t, and the spot wholesale price, p_t:

$$\mu_{q^o}(q, p) = \gamma_5 + \gamma_6 p_t + \gamma_7 p_t^2 + \gamma_8 q_t, \tag{49}$$

and a constant scale parameter, σ_{q^o}.[9] Since in this non-optimizing or reduced form version of the model the evolution of the state vector, (q_t, p_t), does not depend on r, K, ϕ_1, or ϕ_2, these parameters are not estimated.[10] We estimate the parameters $\theta^{rf} = \{\eta_c, \eta_s, \alpha_0^c, \alpha_0^s, \alpha_1^c, \alpha_1^s, \mu_q, \sigma_q, \varsigma, \lambda_p, \mu_p, \sigma_p, \gamma_0, \gamma_1, \ldots, \gamma_7, \gamma_8, \sigma_{q^o}\}$.

5.3. Estimation

We have considerable freedom in our choice of moments functions, the h vector, to use in the criterion. We match the means, variances and histograms (four of the five quintile binds) of the p, p^r, q^o, q^s, and q processes, average markups conditional on the sale size, the fraction of days a purchase is made, the fraction of days a sale is made, and the mean and variance of a 4-day moving average of purchase prices for a total of 42 moment conditions. We set the number of simulations, S, to 100. The Newey–West estimator of $\Omega(h)$ in Eq. (32) requires a bandwidth parameter (number of lags l to estimate autocovariances) to form a consistent estimator $\hat{\Omega}(h)$ of $\Omega(h)$ as $T \to \infty$. Consistent estimation requires the number of lags to be a function of T, $l(T)$, and satisfy $l(T) \to \infty$ as $T \to \infty$ and

$$\lim_{T\to\infty} \frac{l(T)}{T^{1/4}} = 0. \tag{50}$$

We used the bandwidth $l(T) = T^{1/5}$ and for our sample size of $T = 1647$ this implies that $l(T)$ should be either 4 or 5. We experimented with these values and several larger and smaller values. The autocovariances of the MSM residuals die out rather quickly so the results with $l = 5$ or more lags are very similar to results with $l = 4$. However reducing l below 4 affects the estimated standard errors of the MSM significantly and with $l = 0$ (i.e. when $\hat{\Omega}(h)$ is the standard variance covariance matrix estimator that is consistent for *IID* data) the MSM standard errors using formula (41) appeared to be significantly underestimated. Thus, the results we present below are for $l(T) = 4$.

9 We estimated various specifications of Eq. (49) allowing for structural breaks; these alternative specifications did not improve the model's fit of the data significantly.

10 In the profit decomposition exercises presented below, we will use the point estimates from the profit maximizing model for these four parameters to compute and discount profits.

Table 1
Point estimates and standard errors for 3/4 and 1 inch plate.

Parameter	3/4 inch plate		1 inch plate	
	Profit-maximization	Reduced form	Profit-maximization	Reduced form
r	0.0800		0.0758	
	(0.325)		(0.259)	
η_c	−5.907	−5.753	−5.504	−5.534
	(0.067)	(0.233)	(0.092)	(0.395)
η_s	0.326	0.318	0.319	0.305
	(0.069)	(0.015)	(0.095)	(0.021)
α_0^c	−1.159	−1.264	−1.142	−1.011
	(0.0003)	(1.82)	(0.002)	(2.389)
α_0^s	-3.43×10^{-5}	-2.93×10^{-5}	-3.14×10^{-5}	-3.23×10^{-5}
	(0.0036)	(0.0016)	(0.0002)	(0.0003)
α_1^c	0.0752	0.1077	0.0877	0.0883
	(8.42×10^{-5})	(0.026)	(0.0008)	(0.0400)
α_1^s	-4.12×10^{-5}	-4.81×10^{-5}	-4.91×10^{-5}	-4.91×10^{-5}
	(0.0004)	(5.73×10^{-5})	(5.92×10^{-5})	(5.536×10^{-5})
μ_q	9.127	9.824	8.938	9.954
	(0.01)	(0.087)	(0.007)	(0.1212)
σ_q	0.958	0.775	1.212	0.849
	(0.012)	(0.016)	(0.017)	(0.018)
ς	1.561	1.683	1.490	1.658
	(0.016)	(0.0292)	(0.014)	(0.038)
K	4.480		4.401	
	(0.588)		(0.314)	
φ_1	0.00675		0.00593	
	(0.344)		(0.2408)	
φ_2	2.19×10^{-8}		1.44×10^{-8}	
	(3.66)		(1.20)	
λ_p	0.9790	0.9681	0.9782	0.9682
	(5.3×10^{-5})	(0.0002)	(8.44×10^{-5})	(0.0003)
μ_p	0.0594	0.0895	0.0606	0.0900
	(0.0009)	(0.0006)	(0.0027)	(0.0009)
σ_p	0.0214	0.0279	0.0237	0.0305
	(0.016)	(0.0008)	(0.0175)	(0.0012)
γ_0		−0.00797		−0.00810
		(5.04×10^{-12})		(6.21×10^{-12})
γ_1		3.44×10^{-5}		3.19×10^{-5}
		(3.06×10^{-6})		(9.07×10^{-6})
γ_2		−0.0018		−0.00202
		(0.0011)		(0.0005)
γ_3		0.00687		0.00646
		(0.00012)		(0.00023)
γ_4		−3.17		−3.04
		(0.27)		(0.39)
γ_5		7.38		11.81
		(0.042)		(0.15)
γ_6		−0.115		−0.577
		(0.002)		(0.006)
γ_7		0.0030		0.0149
		(6.84×10^{-5})		(0.0003)
γ_8		-5.09×10^{-5}		-1.16×10^{-5}
		(2.44×10^{-6})		(8.63×10^{-7})
σ_{q^o}		0.287		0.255
		(0.012)		(0.006)

We now estimate the profit maximization model and the reduced form model for two products independently. In Table 1 we report the point estimates and standard errors for the parameters for each model for two large construction grade steel plates: one is 3/4 inch thick, the other is 1 inch thick. Although we estimated the parameters for each of these products independently, it is reassuring that the point estimates of the common parameters, r, η_c, η_s, K, α_0^c, α_0^s, α_1^c, α_1^s, λ_p, ς, ϕ_1 and ϕ_2, are similar across the two products and the two behavioral assumptions.

The parameter estimates for the profit maximization model are presented in the second and fourth columns of Table 1. Our point estimates of the annual interest rate r, 8.0% for the 3/4 inch plate and 7.6% for the 1 inch plate, align with our expectations. The general manager would not provide us with specific data on the firm's borrowing and lending (many sales involve trade credit), but told us that one and three-quarter points over a short-term LIBOR rate was a good estimate of the interest rate they faced. The average 3-month LIBOR rate over the period studied is about 5.75%, which implies an average annual borrowing rate for the firm of about 7.5%.

Our estimate of the fixed cost, K, is just a little over \$4 per order. We asked the general manager for his estimate of K. He stated that the cost of placing an order was the value of the time it took for the general manager and his administrative assistant to complete the paperwork; he pointed out that the firm was paying for their time whether an order was made or not.

The demand side parameters are pinned down by the frequency and distribution of sales. We estimate these parameters under both behavioral assumptions. In all four models, the point estimates of $\alpha_0^c, \alpha_0^s, \alpha_1^c, \alpha_1^s$, imply that the customer arrival probability, $1-\eta$, of roughly 0.6 matches the share of days that a sale occurs. The estimates of ς are roughly 1.5 imply that demand is elastic.

Finally, we report the point estimates and standard errors of the coefficients of the reduced form purchasing rule for the two products. The estimated coefficients for the logistic function, P, given in Eq. (48), indicate that the likelihood of a purchase is a decreasing function of inventories on hand ($\gamma_1 > 0$) and the spot price ($\gamma_2 < 0$, $\gamma_3 > 0$), for values of the spot price observed in the data. The firm is highly likely to make a purchase when inventories are at near stockout levels ($\gamma_4 < 0$). The estimated coefficients for the specification of μ_{q^o}, Eq. (49), imply that the size of the purchase will tend to be decreasing in both the spot price ($\gamma_6 < 0$ and $\gamma_7 > 0$) for values of the spot price observed and the quantity of inventories on hand ($\gamma_8 < 0$). However, these coefficients are small and statistically insignificant suggesting that there is only a weak correlation between the state variables and the purchase size.

For the 3/4 inch plate, the purchase decisions observed in the data and predicted by the model under both the profit maximization and the reduced form decision rules are displayed in the three graphs in Fig. 3. In graph (a) we scatterplot in large black dots the (p_t, q_t) pairs for dates that we observe the firm making a purchase. The length of the vertical line segment rising northward from the dot is the purchase quantity, q_t^o, so the line depicts how the purchase propels the firm through the state space. In graphs (b) and (c), we scatterplot a set of simulated state space pairs (p_t, q_t) for the model under the two behavioral assumptions. Analogous to the data plotted in graph (a), we denote the (p_t, q_t) pairs on the days which the model dictates a purchase with large black dots and draw vertical line segments of length q_t^o originating from each dot. To highlight the endogenous sampling problem, in both graphs, we scatterplot small blue circles denoting the (p_t, q_t) pairs from the simulation for days in which no purchase was made.

Consider graph (b). According to the optimal trading rule, the firm only makes purchases when the (p_t, q_t) pair is on or below the $s(p)$ band. Thus the black dots align along this downward sloping curve in the southwest corner of the graph. That is, the firm only purchases when prices are low and inventories are low. As the spot price increases, the threshold quantity of inventory under which the firm makes a purchase decreases. In the simulation presented, purchases occur less than 15 percent of time. When the firm decides to make a purchase, the purchase quantity (the length of the vertical line) is given by $q_t^o = S(p) - q_t$. Due to the fixed costs of ordering, the $S(p)$ band is strictly above the $s(p)$ band although the difference between the two bands decreases as the price increases. In other words, the minimum order size is a decreasing function of the price. The "buy low" strategy is not driven by the fixed cost, K, which we estimated to be less than \$5.00 per transaction.

Now consider graph (c). In the model with the randomized decision rule, the cloud of large black dots, denoting (p_t, q_t) pairs on which a purchase occurred, is concentrated on the western side of the state space and is downward sloping. In contrast the small blue circles, denoting the simulated (p_t, q_t) pairs on which no purchase occurred, are more evenly distributed across the state space. Thus, under this reduced form rule, the firm is more likely to purchase steel when prices are low than when prices are high. As suggested by the small size of the coefficients on Eq. (49), the purchase size is weakly correlated with the spot price or current inventories. In contrast to the theoretical predictions of both the profit maximizing model and the model with randomized decision rules, the cloud of black dots denoting purchases in graph (a) is generally upward-sloping indicating that neither the likelihood of purchases nor the size of the purchase is negatively related to the current spot or the firm's current level of inventories.

Finally, note that in graph (a) we observe roughly a dozen large purchases in the data. In graph (c) we see that the reduced form model replicates these infrequent but large purchases; while in contrast, in graph (b), under the profit maximization assumption, there are no large purchases. The ability of the reduced form model to accurately replicate the distribution of purchases is confirmed in the densities of purchases plotted in Fig. 4. These densities confirm that the profit maximization model overestimates the frequency of small purchases while underestimating the frequency of large purchases for both products. The density of purchases for the reduced form model aligns with the shape of the density of purchases from the data.

In Table 2 we report a subset of the 42 moments we used in the MSM estimator along with the average moments from 100 simulations of the models. In general, for both types of plates, the reduced form model does a better job matching the data moments than the profit maximization model. For both types of plates, the minimized MSM estimation criterion, reported in the bottom row of the table, is lower for the reduced form model than for the profit maximization model.

As in the data, in all four models the firm makes a purchase on 14 percent of days and makes a sale on roughly 60 percent of days. The models also correctly replicate the fact that the average purchase size is four times the size of the average sale. But, as can be anticipated by the discussion above, the profit maximization model underestimates the mean and the variance of the purchase size, while the reduced form model does a better job matching the levels of these two moments; as a consequence, the reduced form model also more closely matches the mean and variance of inventory holdings.

Turning to prices, all four models underestimate the mean and variance of the buy and sell prices. The models succeed in replicating the negative relationship between the markup and sales quantity; as in the data, the larger is the quantity

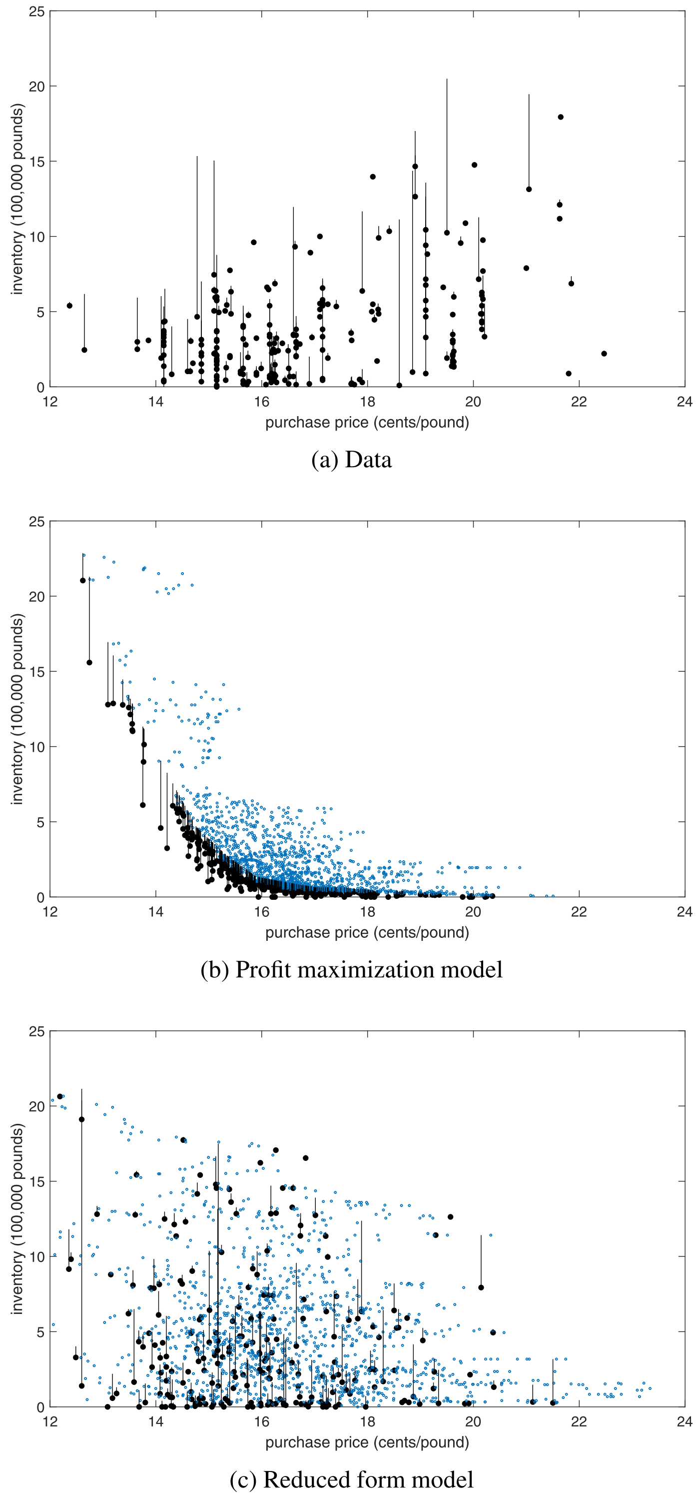
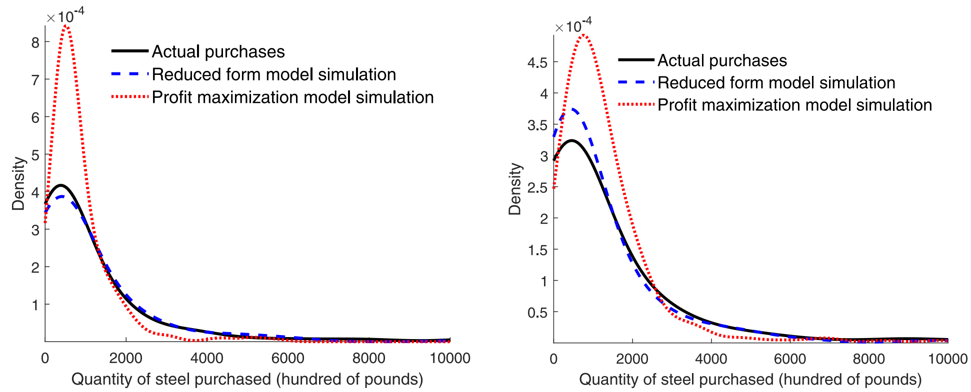


Fig. 3. Scatterplots of purchase prices and inventories for 3/4 inch plate.

sold, the smaller is the markup. This pattern holds in all four models though the models underestimate the size of the markup. The average markups in the reduced form model are closer to the average markups in the data.

The main shortcoming of these models is our inability to match the downward trend of the price process that we see in almost all of the firm's products. As illustrated in Fig. 1 the wholesale price for the 3/4 inch plate fell from 20 cents per pound in 1997 to about 12 cents per pound in 2002. In all four models, prices are stationary though highly persistent. Consequently, as can been seen in graph (b) of Fig. 3 the optimal decision rules under profit maximization imply counterfactually that the firm should make only small purchases and hold low levels of inventories whenever the



(a) 3/4 inch plate (b) 1 inch plate

Fig. 4. Densities of steel purchases.

Table 2
Comparison of selected moments from 100 simulations and data for 3/4 inch and 1 inch plate.

Moment	3/4 inch plate			1 inch plate		
	Profit max model	Reduced form model	Data	Profit max Model	Reduced form model	Data
Mean (buy price)	16.76	16.06	16.78	15.89	16.35	16.95
Var (buy price)	3.32	2.84	4.37	3.39	3.44	5.58
Mean (sell price)	18.05	18.18	19.09	17.45	18.15	18.96
Var (sell price)	3.00	3.61	5.37	3.40	4.26	5.58
Mean (purchase size)	786	982	1138	1105	1237	1393
Var (purchases)	77	240	370	162	410	629
Mean (sale size)	180	230	260	251	291	313
Var (sales)	5	4	5	18	9	12
Mean (inventory)	3943	4937	5247	6091	7500	7561
Var (inventory)	2369	1385	1589	4505	3635	4067
Mean (markup) | small sale	1.55	2.02	2.14	1.71	1.79	2.02
Mean (markup) | medium sale	1.41	1.88	1.71	1.57	1.65	1.79
Mean (markup) | large sale	1.14	1.60	1.63	1.17	1.31	1.59
Share of days with buys	0.143	0.143	0.144	0.138	0.139	0.140
Share of days with sales	0.587	0.601	0.605	0.572	0.577	0.594
Minimized criterion	1.05	0.17		0.89	0.14	

Prices are in cents per pound. Quantities are in hundred weight (i.e. 100s of pounds).

procurement price is above 17 cents per pound. From Fig. 2 we see that, for the 3/4 inch plate, the firm made large purchases around 18.5 cents per pound in April 1998, and around 15 dollars per hundred-weight in the later part of the sample.

An often suggested solution to this trend problem is that we assume prices follow a random walk. However if we assume the price process follows a (or a very nearly) random walk, the optimal decision rules imply frequent small- to medium-size orders such that the inventory level fluctuates closely around a fixed target level. A version of the model which assumes p_t follows a random walk will not imply the large variation in inventory holdings that we see in the data. A second potential solution is to detrend the data. However when we first started working on this project, no one we talked to expect steel prices to decline 40% in four years. To some extent we are just working with too short a sample period. A third candidate solution is to add an additional macroeconomic state variable. Such a variable could allow for "high price" regimes and "low price" regimes. We view this third solution as the most promising.

Finally, since the profit maximization model is a limiting special case of the unrestricted reduced form model, we can use the Hausman (1978) specification test to evaluate the consistency of the restrictions imposed by profit maximization. For the twelve parameters in common across the two specifications, reported in Table 1, the results of the Hausman test reject the null hypothesis that the profit maximization model is correctly specified. The test statistics for the 3/4 inch plate and 1 inch plate are 3688 and 1458, respectively. Both statistics are distributed chi-squared with 12 degrees of freedom. Thus, we can reject the assumption that the firm is behaving optimally.

5.4. A profit decomposition exercise

Finally, we use simulations of the four estimated models to deduce the relative importance of capital gains versus markups for the overall profitability of the firm. By substituting the law of motion for inventories, (4), into the firm's

Table 3
Profit decomposition for 3/4 inch and 1 inch plate using equation (51).

3/4 inch plate	G.M.'s actual Performance			Profit Maximization decision rules			Reduced form model's decision rules		
Markup	\$400,464	(13,237)	90%	\$291,444	(13,677)	58%	\$367,032	(16,187))	100%
Capital gain	45,770	(13,945)	10%	210,045	(13,528)	42%	−362	(29,460)	0%
Holding cost	−52,446	(0)		−84,892	(4,213)		−59,750	(11,342)	
Order costs	−885	(0)		−1032	(42)		−855	(59)	
Total profits	392,903	(1,522)		415,564	(15,140)		306,065	(30,229)	
1 inch plate	**G.M.'s actual Performance**			**Profit maximization decision rules**			**Reduced form model's decision rules**		
Markup	\$446,854	(17,646)	86%	\$328,010	(15,718)	47%	\$410,146	(20,373)	97%
Capital gain	75,222	(18,694)	14%	369,942	(23,736)	53%	14,118	(38,306)	3%
Holding cost	−65,895	(0)		−118,341	(9,111)		−52,403	(9,600)	
Order costs	−834	(0)		−1018	(51)		−926	(59)	
Total profits	455,348	(2,233)		578,593	(24,330)		370,935	(39,086)	

Both the actual and the counter-factual profits cover the 1647 days studied and are discounted back to the start of the sample period, July 1, 1997. The profit numbers reported are the average across 100 simulations. The numbers in parentheses are the standard deviations from the 100 simulations. Total profits are the sum of the first four rows.

objective function, (44), the discounted present value of the firm's profits can be expressed by

$$\sum_{t=1}^{T}\beta^t\pi(p_t,p_t^r,q_t^r,q_t+q_t^o)=\sum_{t=1}^{T}\beta^t(p_t^r-p_t)q_t^s+q_1p_1+\sum_{t=2}^{T}\beta^t(p_t-(1+r)p_{t-1})q_t-\sum_{t=1}^{T}\beta^tI(q_t^o)K-\sum_{t=1}^{T}\beta^tc^h(q_t+q_t^o,p_t). \tag{51}$$

The first term on the right hand side of Eq. (51) can be interpreted as the discounted present value of the markup paid by the firm's retail customers over the current wholesale price while the third term can be interpreted as the discounted present value of the capital gains or losses from holding the steel from period $t-1$ into period t. The fourth and fifth terms are the discounted present values of the order costs and the holding costs incurred by the firm over the sample period.

Since this decomposition depends on the wholesale price path between purchases, we simulate between purchase dates via importance sampling. That is, for each interval between successive purchase dates, we simulate wholesale price paths that are consistent with the estimated law of motion (45) and the observed purchase prices at the beginning and end of the interval. Since our theory implies that the firm places an order anytime the quantity falls below the order threshold, $s(p)$, we truncate the simulated price process by rejecting any paths such that $q_t < s(p_t)$ for any draw within the simulated realizations using an approach developed by Keane (2009).

We first employ this decomposition to evaluate the general manager's actual performance over the six-and-a-half year sample period for the two plates. For a given interpolated price series, we decomposed the firm's profits using the actual data for q_t, q_t^s, and q_t^o and our point estimates for r, K, ϕ_1, and ϕ_2. In Table 3 we report the average decomposition from 100 simulated wholesale price paths. As discussed in the introduction, the price of steel fell steadily over the sample period. Even so, by our accounting, the firm made \$392,903 (3/4 inch plate) and \$455,348 (1 inch plate) from the markup and capital gains net holding and order costs on each of these two products over the six-and-a-half year period.[11] Ignoring the fixed order and holding costs, about 90 percent (3/4 inch plate) and 86 percent (1 inch plate) of these profits came from the markup, while the remaining 10 and 14 percent came from capital gains. We find it remarkable and evidence of the general manager's acumen in steel trading that the firm made positive capital gains over this period despite the price of steel falling about 40 percent. While the firm's success in price speculating is good for its profits, it increases the potential biases from failing to account for the endogeneity of the sampling process.

As a diagnostic of our theory, we compare the general manager's performance to the profit maximization and reduced form models' predictions. In this exercise we take as given the 100 interpolated wholesale price series, the firm's quantity demanded series, and the firm's initial level of inventories for each product. But in this case, we let the model's decision rule (i.e., the optimal $(S(p), s(p))$ rule from the profit maximization model or the randomized (P, F) rule from the reduced form model) dictate when and how much to order. Inventories follow the accumulation identity given by Eq. (4). As reported in Table 3, had the general manager counter-factually followed the optimal order strategy implied by our model, his discounted profits from the markup would have been considerably smaller: \$109,020 less for 3/4 inch plate; \$118,844 less for 1 inch plate. However, his capital gains would have been considerable larger: \$164,275 more for 3/4 inch plate; \$294,720 more for 1 inch plate.

11 Profits are discounted back to July 1, 1997.

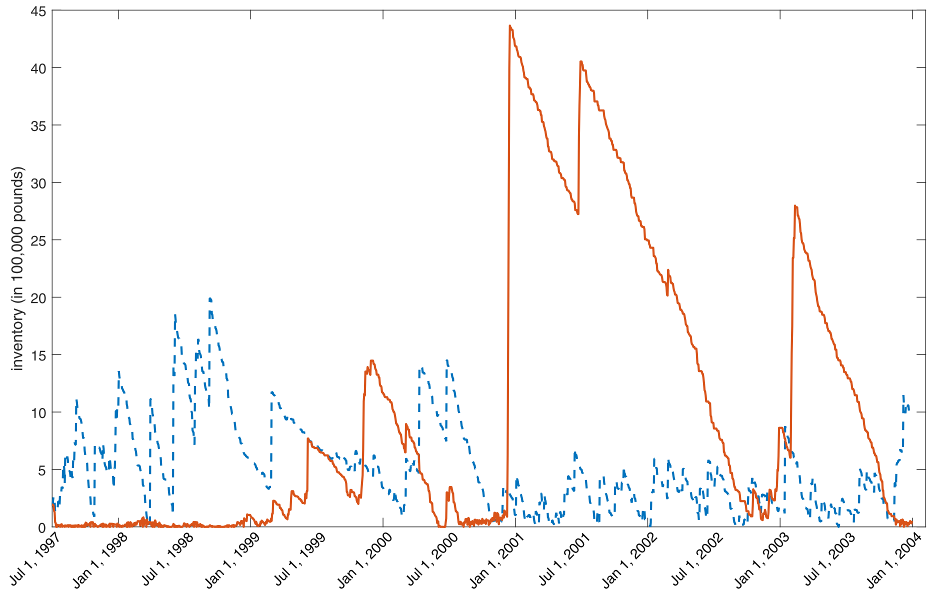


Fig. 5. Actual (dashed line) and counter-factual under profit maximization (solid line) inventory holdings for 3/4 inch plate.

The (S, s) model implies that the firm should engage in aggressive price speculation. Under the (S, s) strategy, the firm earns roughly half its profits from capital gains (i.e., buying low and selling high). In contrast the purchasing decisions from the unrestricted, reduced form model is consistent with the firm's actual performance and suggests that the firm should earn the lion's share of its profits from the markup. The near-zero average capital gain from the reduced form model is mainly due to the downward trend in prices over the sample period. Despite the reduced-form model's ability to match the distribution of general manager's purchases, following the predictions of this model would have led to roughly $80,000 less in profits per product relative the G.M.'s actual decisions. Hence, while the G.M. was not purely profit maximizing, his decision making led to higher profits both from markups and capital gains than can be replicated from a simple reduced-form trading rule.

In Figs. 5 and 6 we plot the prices and inventory holdings for one simulation of the profit maximization model. In Fig. 5 we plot both the actual inventory holdings along with the implied holdings under the model's optimal decision rules. In Fig. 6 we plot the corresponding retail and wholesale price paths. The model's counter-factual inventory path differs considerably from the firm's actual inventory path. In the beginning of the sample, years 1997 and 1998, when prices were high, the model implies the firm should have made frequent small purchases and held relatively low levels of inventories. As was discussed in the introduction, in April 1998 when the wholesale price of steel dropped from 20 cents per pound to 18.5 cents per pound, the firm built up its inventory of 3/4 inch plate substantially. In contrast the model does not view 18.5 cents as a particularly good price; as can be seen in the (S, s) bands plotted in Fig. 3, the target inventory level at 18.5 cents is around 100,000 pounds. In April 1998, the firm's inventory of 3/4 inch plate was roughly 2,000,000 pounds.

It is not until December 1999 when prices fell below 13 cents a pound that the model recommends holding more than 1,000,000 pounds of inventory. However during January and February 2000, the general manager let his inventory of 3/4 inch plate fall to almost zero. The sharp contrast between the model's counter-factual inventory policy and the firm's behavior is also evident during the second half of the sample. In mid-December 2000, the general manager had an opportunity to buy steel for a little over 10 cents per pound The (S, s) model would have purchased very large quantities at these prices.[12] In years 2001 and 2002, the firm held relatively low levels of inventories, whereas the (S, s) model took the opportunity to "stock up" resulting in inventory levels in excess of 2,000,000 pounds. Essentially the (S, s) model predicts purchasing behavior over this period that is nearly the opposite of what the firm actually did: the firm held low inventory levels in 1997, 1998 and 1999, and high inventory levels from 2001 to the start of 2004, almost the reverse of what the (S, s) strategy would have done.

This counter-factual exercise is rigged in the profit maximization model's favor in one dimension and rigged against the model in another. Since we used the entire sample period to both estimate the model and evaluate the model's performance, the model "knows" the mean and the standard deviations of prices and quantity demanded for the entire period. The model knows, whereas the general manager did not know, that a price of 18.5 cents per pound in the Spring

12 In fairness to the general manager, he bought as much steel as he could at these prices. If we constrained the model to purchase no more steel than he actually did on those particular days, Fig. 5 would not be quite as dramatic; but the model would still imply that inventories should considerable higher in years 2001 and 2002 than the levels we observed in the company's records.

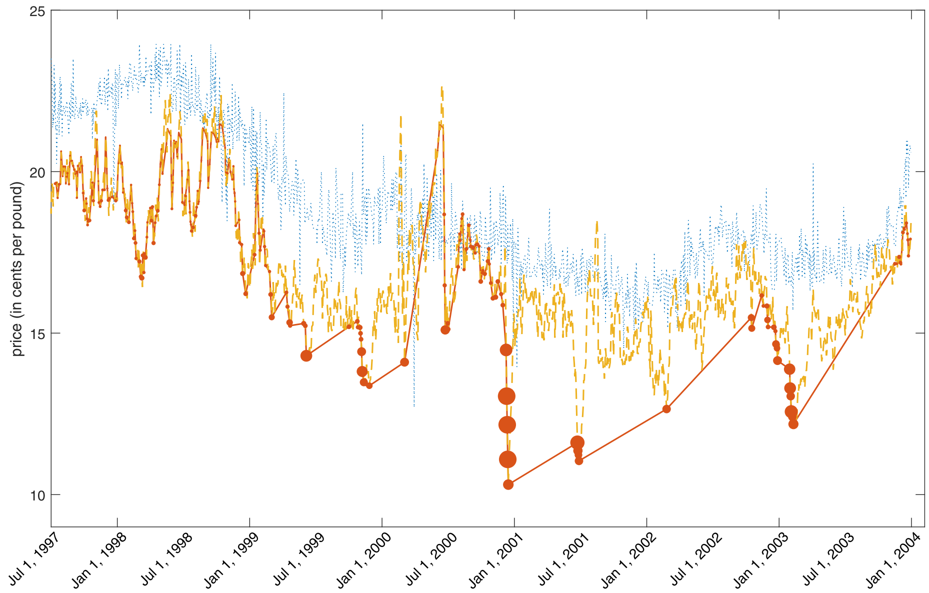


Fig. 6. Counter-factual uncensored purchase prices (dashed line), censored purchase prices under profit maximization (solid line), and retail prices (dotted line) for 3/4 inch plate. For the censored purchase price series, the size of the marker is proportional to the size of the purchase.

of 1998 was an above-average price for the 1997–2003 period. In this way the model has an advantage over the manager. However the model is constrained to sell at most the quantity of steel that the general manager actually sold. The model does not get the opportunity make any sales the general manager might have had the option to make but decided to turn down.

While we do not report an out-of-sample comparison between our model and the general manager, if we had estimated the model through the Fall of 2001, and then used our model to dictate purchases for the firm for the Winter and Spring of 2002, our model would have outperformed the general manager. In the Fall of 2001, the firm was purchasing steel for around 10 to 12 cents per pound. We told the general manager at that time that our model recommended building up inventories at these prices. He did not follow this advice since he anticipated further price declines. He argued (and to be honest, we did not disagree) that our model did not take into account the potential slowdown in the economy in the wake of the terrorist attack of September 11, 2001 and possible resulting reduction in the demand for steel. He also expected new production capacity from the Nucor Corporation to put additional downward pressure on prices. However, with the bankruptcy of Bethlehem Steel in October 2001 as well as both the anticipation of an increase and the actual increase in steel tariffs imposed by President Bush in March 2002, steel prices increased about 20 percent in the Spring of 2002 to the 12 to 14 cent range. In the Spring of 2002, we reminded the general manager that in the fall our model recommended he build up inventories. He sighed, "I wish I had."

In this case, our model "got it right" but perhaps not for the right reasons. Our model was predicting an increase in prices since our model always expects prices to return to the sample mean. Our model had no way of predicting where the economy was going. It just expected demand to be stationary; moreover, our model could provide no help in forecasting the President's actions regarding steel tariffs. In future work, we hope to add a macroeconomic state variable to the analysis. Thus we could use our model jointly with a macroeconomic forecasting model to provide conditional inventory level recommendations to the firm such as "If you expect the economy to remain strong, the model recommends holding inventories in a range from X to Y; if you expect the economy to weaken, ..."

6. Conclusion

A huge and highly influential econometric literature has been developed to enable us to do valid inference in the presence of various types of censoring, or more generally *endogenous sampling* including the work of Heckman (1979) on "sample selection bias" and the work of Manski and McFadden (1981) on "choice based sampling." Most of the early work has been done in cross sectional contexts, but there is also a growing time series literature that recognizes and deals with the econometric problems arising from irregular sampling of financial time series (Engle and Russell, 1998).

We analyzed a problem where there is both irregular and endogenous sampling of time series data on prices by a firm that operates as a speculator in the steel market. Since the firm only records the wholesale prices of steel on the days it buys steel, we have no data on prices on the vast majority of days where it does not buy steel. We need to correct for this endogeneity to make valid inferences. For example, since the firm profits from "buying low and selling high" the average price of steel on the days the firm buys steel is a downward biased estimate of the ergodic or long run average wholesale price of steel.

We demonstrated that it is possible to use McFadden's (1989) MSM estimator to deal with the endogenous sampling problem and consistently estimate the parameters of an AR(1) specification of the wholesale price of particular steel products. The MSM estimator finds parameters that make a vector of actual moments calculated from the data as close as possible to a vector of simulated moments from the model. The MSM estimator accommodates endogenous sampling in an extremely intuitive and straightforward manner: we simply sample the simulated data in the same way the actual data are sampled. Namely, in both the simulated and actual datasets we are only allowed to observe wholesale prices on days where the firm purchases steel. The immense flexibility and ease by which MSM can deal with a variety of endogeneity and missing data problems has made it the equivalent of the "Swiss Army knife" in applied econometrics, especially for structural estimation of models which are difficult or impossible to estimate via other means such as maximum likelihood.

In this paper we estimated two specifications of a model of speculative steel trading by MSM. One model imposed the hypothesis of optimality, and we showed that the optimal speculative trading strategy takes the form of a generalized (S, s) rule. This structural version of the MSM estimator requires a nested numerical solution of the firm's dynamic programming problem to compute the (S, s) threshold functions. We then used the numerically calculated (S, s) bands to simulate purchases, sales, and inventory holdings for this firm. The other unrestricted model specification relaxes the assumption of optimality and allows the firm's purchases to be given by an arbitrary, potentially randomized Markovian purchasing strategy. We used a flexible sieve approximation to this randomized strategy and also used it to simulate purchases, sales, and inventory holdings. Thus, the unrestricted model relaxes the assumption of optimality but has the flexibility to approximate an optimal deterministic (S, s) strategy arbitrarily closely as the number of terms in a series approximation to the probability distributions encoding the randomized strategy increases.

Our empirical conclusion is that the actual trading strategy is much better approximated by a randomized purchasing strategy than the deterministic (S, s) strategy, and the unrestricted model fits the data much better than the restricted model that imposes the assumption of optimality. We formally tested and rejected the hypothesis of optimality using a Hausman test, and we showed via a simulation study that the counterfactual (S, s) trading strategy would have resulted in higher profits compared to what the firm actually earned over the period of our study, and confirms that the unrestricted randomized strategy that approximates the firm's actual purchasing strategy is indeed suboptimal, earning only between 50% and 66% of the profits achievable under an optimal (S, s) rule.

In future work it would be desirable to try to verify the regularity assumptions underlying the asymptotic normality of the MSM estimators, particularly to verify the conditions on the series estimator of the firm's purchasing strategy (P, F) for the unrestricted "reduced form" specification necessary to establish the asymptotic normality of the parameters of interest. We also think it is of interest to compare other potentially more efficient estimators, including maximum likelihood estimators. As we noted maximum likelihood is computationally challenging due to the high dimensional numerical integrations required. However simulated maximum likelihood (used for example in the study by Erdem et al. (1999) to deal with endogenously sampled price data in a panel data context), or direct maximum likelihood using the recursive likelihood integration (RLI) algorithm of Reich (2018) may be computationally feasible and result in a more efficient estimator that is less subject to the well known problems in optimizing the non-smooth MSM criterion function. However due to the problem of statistical degeneracy of the structural (S, s) model, we would have to extend this model to incorporate econometric unobservables such as unobserved costs of placing orders to implement maximum likelihood under the hypothesis of optimality.

While this research was motivated by a new dataset from a single steel wholesaler, most datasets in which agents have the choice of whether and when to participate in a market activity will be endogenously sampled. In most markets, the only prices recorded are the transaction prices — econometricians almost never get to observe prices offered but not transacted on. For example, econometricians rarely get to observe the wages unemployed job seekers are offered but refuse. It should be straightforward to apply the MSM estimator to other types of endogenous sampling problems that arise in time series contexts.

References

Aguirregabiria, V., 1999. The dynamics of markups and inventories in retailing firms. Rev. Econom. Stud. 66, 275–308.
Ai, C., Chen, X., 2003. Efficient estimation of models with conditional moment restrictions containing unknown functions. Econometrica 71–6, 1795–1843.
Aït-Sahalia, Y., Mykland, P.A., 2001. The Effects of Random and Discrete Sampling when Estimating Continuous-Time Diffusions. Princeton University, working paper.
Allenby, G., McCulloch, R., Rossi, P., 1996. On the value of household information in target marketing. Mark. Sci. 15, 321–340.
Arrow, K.J., Harris, T., Marschak, J., 1951. Optimal inventory policy. Econometrica 19–3, 250–272.
Athreya, R., 2002. Price Dispersion in the Wholesale Market for Steel. Yale University, manuscript.
Chan, H.M., 2001. Analysis of Variance in Steel Sales Price. Yale University, manuscript.
Chen, X., 2007. Large sample sieve estimation of semi-nonparametric models. In: Heckman, J., Leamer, E. (Eds.), Handbook of Econometrics, Vol. 6B. Elsevier.
Chen, X., Shen, X., 1998. Sieve extremum estimates for weakly dependent data. Econometrica 66–2, 289–314.
Deaton, A., Laroque, G., 1992. On the behavior of commodity prices. Rev. Econom. Stud. 59, 1–23.
Duffie, D., Glynn, P., 2004. Estimation of continuous-time Markov processes sampled at random time intervals. Econometrica 72–6, 1773–1808.
Duffie, D., Singleton, K.J., 1993. Simulated moments estimation of Markov models of asset prices. Econometrica 61–4, 929–952.
Engle, R., Russell, J., 1998. Autoregressive conditional duration: A new model for irregularly spaced data. Econometrica 66–5, 1127–1162.
Erdem, T., Keane, M., 1996. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. Mark. Sci. 15–1, 1–20.

Please cite this article as: G. Hall and J. Rust, Estimation of endogenously sampled time series: The case of commodity price speculation in the steel market. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.07.033.

ARTICLE IN PRESS

Erdem, T., Keane, M., Sun, B., 1999. Missing price and coupon availability data in scanner panels: Correcting for the self-selection bias in choice model parameters. J. Econometrics 89, 177–196.
Fabian, T., Fisher, J.L., Sasieni, M.W., Yardeni, A., 1959. Purchasing raw material on a fluctuating market. Oper. Res. 7, 107–122.
Frazier, D., Zhu, D., 2017. Derivative-Based Optimization with a Non-Smooth Simulated Criterion. Department of Econometrics, Monash University, manuscript.
Futia, C., 1982. Invariant distributions and the limiting behavior of Markovian economic models. Econometrica 50–2, 377–408.
Golabi, K., 1985. Optimal inventory policies when ordering prices are random. Oper. Res. 33–3, 575–588.
Grenander, U., 1981. Abstract Inference. In: Wiley Series, New York.
Hall, G., Rust, J., 1999. An empirical model of inventory investment by durable commodity intermediaries. In: Carnegie-Rochester Conference Series on Public Policy, Vol. 52.
Hall, G., Rust, J., 2007. The (S, s) rule is an optimal trading strategy in a class of commodity price speculation problems. Econom. Theory 30, 515–538.
Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1929–1954.
Hausman, J.A., 1978. Specification tests in econometrics. Econometrica 46–6, 1251–1271.
Hayaski, T., Kusuoka, S., 2008. Consistent estimation of covariation under nonsynchronicity. Stat. Inference Stoch. Process. 11, 93–106.
Heckman, J., 1979. Sample selection bias as specification error. Econometrica 47–1, 153–161.
Heckman, J., 1981. Statistical models for discrete panel data. In: Manski, C., McFadden, D. (Eds.), Structural Analysis of Discrete Data. MIT Press.
Kaldor, N., 1939. Speculation and economic stability. Rev. Econom. Stud. 7, 1–27.
Kalymon, B., 1971. Stochastic prices in a single-item inventory purchasing problem. Oper. Res. 19, 1434–1458.
Keane, M., 2009. Simulated maximum likelihood estimation based on first-order conditions. Internat. Econom. Rev. 50–2, 627–675.
Kingman, B.G., 1969. Commodity purchasing. Oper. Res. Q. 20–1, 59–79.
Koenker, R., 1988. Asymptotic theory and econometric practice. J. Appl. Econometrics 3–2, 139–147.
Lee, B., Ingram, B.F., 1991. Simulation estimation of time-series models. J. Econometrics 47, 197–205.
Manski, C., McFadden, D., 1981. Alternative estimators and sample designs for discrete choice analysis. In: Manski, C., McFadden, D. (Eds.), Structural Analysis of Discrete Data. MIT Press.
McFadden, D., 1989. A method of simulated moments for estimation of discrete models without numerical integration. Econometrica 57–5, 995–1026.
McFadden, D., 1997. On the Analysis of Endogenously Recruited Panels. University of California at Berkeley, manuscript.
Miranda, M., Rui, X., 1999. An Empirical Reassessment of the Nonlinear Rational Expectations Commodity Storage Model. Ohio State University, manuscript.
Moinzadeh, K., 1997. Replenishment and stocking policies for inventory systems with random deal offerings. Manage. Sci. 43–3, 334–342.
Newey, W., Powell, J., 2003. Instrumental variables estimation of nonparametric models. Econometrica 71, 1551–1569.
Newey, W., West, K., 1987. A simple, positive definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.
Özekici, S., Parlar, M., 1999. Inventory models with unreliable suppliersin a random environment. Ann. Oper. Res. 91 (0), 123–136.
Potiron, Y., Mykland, P.A., 2017. Estimation of integrated quadratic covariation with endogenous sampling times. J. Econometrics 197–1, 20–41.
Reich, G., 2018. Divide and conquer: Recursive likelihood function integration for hidden Markov models with continuous latent variables. Oper. Res. 66–6, 1457–1470.
Rust, J., Hall, G., 2003. Middlemen versus market makers: A theory of competitive exchange. J. Polit. Economy 111–2, 353–403.
Scarf, H., 1959. The optimality of (S, s) policies in the dynamic inventory problem. In: Arrow, K., Karlin, S., Suppes, P. (Eds.), Mathematical Methods in the Social Sciences. Stanford University Press, Stanford, CA.
Song, J., Zipkin, P., 1993. Inventory control in a fluctuating demand environment. Oper. Res. 41–2, 351–370.
Vázquez-Abad, F.J., Kushner, H.J., 1992. Estimation of the derivative of a stationary measure with respect to a control parameter. J. Appl. Probab. 29–2, 343–352.
Williams, J.C., Wright, B., 1991. Storage and Commodity Markets. Cambridge University Press, New York.

Please cite this article as: G. Hall and J. Rust, Estimation of endogenously sampled time series: The case of commodity price speculation in the steel market. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.07.033.