Production Decisions During Shortages*

Siddhartha Lewis-Hayre

January 2025

Work in Progress
Click here for the latest version.

Abstract

How efficiently do firms use homogeneous inputs during shortages? I extend Lewis-Hayre (2024) to capture this decision, and I derive an identification strategy to measure the variation in the marginal value of inputs between firms. I apply the test to the supply chain disruptions in the automobile industry during COVID. I show that auto-manufacturers' production followed a pecking order defined by prioritizing the production of car models with the greatest markup in dollar terms. In ongoing work, I compute the optimal inventory levels of automobile parts under the counterfactual assumption of an efficient secondary market during shortages. As another result of the pecking order, shortages are more severe for cars with lower markups. In ongoing work, I show how this inhibits access to new cars for lower income consumers during shortages.

*I thank Professor Adair Morse for her guidance.

1 Introduction

Particularly since the supply chain crises of the COVID era, governments have deemed access to critical inputs a matter of national security. Policy has reflected this stance through legislation like the CHIPS and Science Act. One particularly intriguing solution is a nascent plan by the Biden Administration to create a sovereign wealth fund which could potentially buy futures to secure access to commodities in times of stress. On one hand, such industrial policy aims to remove frictions from technological development. These policies also seek to secure access to critical inputs in times of stress. Why should the government intervene in securing critical inputs?

One reason could be that firms do not use inputs efficiently during shortages. This would distort firms investments in supply chain resilience. I show in Lewis-Hayre (2024) that uncertainty and supply frictions associated with crises can cause a misallocation of inputs between firms. In this paper, I extend the model to the case of producers. I derive an identification strategy to measure the heterogeneity in the marginal value of an input across producers. This allows me to quantify the misallocation of goods during shortages.

How can a researcher measure a firms' marginal value of an input? Consider a producer that produces multiple product varieties that share homogeneous inputs. To maximize profits during shortages, the producer should allocate the input to the variety with the highest markups in dollar terms. As such the producer follows a production pecking order. In practice, however, it is costly to adjust production decisions. Frictions in adjusting production decisions mean that firms will not perfectly follow this pecking-order. To what extent do producers follow this pecking order and reallocate inputs to their most profitable goods when facing input shortages? Answering this question will serve as an identification strategy to understand the cost of adjusting production decisions, which will uncover the value of an extra input to different manufacturers.

I test the pecking order using the auto industry. The automobile industry is of particular interest because automobiles are an essential consumer good whose production relies heavily on semiconductors, which are central to policy on securing critical inputs for crises. The automobile provides a way to test the theory since automobile manufacturers produce multiple models of cars and because different models use some of the same parts.

For empirical identification, I study the supply shortages of 2021, during the COVID pandemic. Starting in February 2021, inventory levels plunged even more than they had at the onset of COVID in March 2020. Figure 1 shows inventory overtime broken into high, medium, and low priced cars.

During this period of February 2021 to March 2022, inventory of more expensive cars fell less sharply. As a result, the average price of a vehicle in inventory shot up during 2021 despite a relatively flat trajectory before and after that year as shown in Figure 2. Figure 3 shows that at least part of the reason for the increase in the average price of a car in inventory was that production of the less expensive cars fell by more than that of expensive cars relative to pre-COVID levels. Why did higher priced cars' production fall by less during inventory shortages?

One explanation could be producers' expectation of demand. Perhaps strain on middle-income households' balance sheets during COVID caused a reduction in demand for less expensive cars and producers responded by lowering production for those cars. Another explanation could be that supply chain disruptions were more severe for cars with a lower price. Extending Grossman, Helpman and Lhuillier (2023) to consider goods with different market share functions would show that producers should more heavily invest in the resilience of the supply chains of goods with lower demand elasticity. For this reason, perhaps supply chains are more resilient for cars with the largest markups in dollar terms, which also likely have higher prices.

Another possibility is that producers followed the pecking order and reallocated inputs to produce their most profitable goods. Such a reallocation could arise either from supply shocks or incorrect producer expectations for demand. The subsequent analysis is agnostic on whether demand or supply caused the shock. An identification strategy to test the reallocation argument will have to show that the two alternative explanations of the data are not driving the results. To show that producers' expectations for demand does not explain the results, I look for strategies to infer producers' expectations for demand. To confront the possibility that producers invest more in supply chain resilience for profitable cars, I need to level cars on the degree to which supply chain disruptions impacted their production. I propose robustness tests to deal with these two possibilities.

Even if manufacturers reallocate parts between cars, at what stage of the production process does this reallocation happen? Four potential levels for input reallocation are: producer, plant, platform, and model. At the producer level, producers may truck delivered shipments of inputs to factories that make the more profitable cars. At the plant level, producers may strategically use the parts delivered to the plant to build the most profitable cars. At the platform level, production lines may prioritize profitable models. At the model level, producers may focus on different trims when faced with shortages. For now I only test the producer and plant level, the other two levels are ongoing research. I test the pecking order at each of these four levels to understand producers'

use of scarce inputs in times of crisis.

1.1 Methodology

The empirical methodology builds off an identification strategy derived from Lewis-Hayre (2024). The details of the model and the derivation of the empirical strategy are in the appendix.

1.1.1 Critical Input Specification

For starters, I simplify the full empirical strategy and assume that every car requires a single critical input. In other words, every producer can make a certain number of cars each period, and the producer decides how to allocate this production budget across cars. The empirical specification simplifies to:

$$\hat{y}_{ijt} = \beta_1 p_{ijt}\sigma_{ijt} + \beta_2\sigma_{ijt} + \beta_3 p_{ijt} + \gamma_{it}\chi_{it} + \gamma_{ij} + \epsilon_{ijt} \tag{1}$$

Where $\hat{y}_{ijt}$ is the production of firm i of product j at time t as a percent of the optimal level of production. I refer to it as the "Production (%)". Additionally p_{ijt} is the price of product j at time t, and σ_{ijt} is producer i's expected ability to sell the marginal unit of product j at time t. Furthermore, $\chi_{it} = \mathbb{1}_{\iota_{it+1}=0}$ is an indicator specifying whether producer i faces a shortage of parts at time t, and ϵ_{ijt} is the error term. The specification has model fixed effects, γ_{ij}, which represent the marginal cost of the car; input fixed effects, γ_k, which represent the cost of an input; and producer-time fixed effects, γ_{ik}, which are only activated when producer i faces a shortage of parts. These scarcity fixed effects represent the shadow cost (above the normal cost) of the critical input at time t to producer i. The coefficient of interest is β_1, and the null hypothesis is that $\beta_1 = 0$.

This specification is an extreme case where the only part in short supply is an identical part, perhaps a semiconductor, that is used in every car. The specification asks: when shortages are worse, do manufacturers focus more heavily on production of cars with a higher price?

The control σ_{ijt} is a first step at leveling cars on how severely their supply chains were disrupted. The control p_{ijt} absorbs the effect of the price of a model changing when a new generation is released.

In this specification, I use inventory as a proxy for the expected number of months to sell the marginal car. In particular, σ_{ijt} is inventory of car j in time t normalized by a baseline inventory level. I call this variable "Inventory (%)". Without normalization, this variable would not be comparable across cars since producers' optimal level of months inventory varies significantly

between cars. The assumption behind this normalization is that the optimal level of inventory for each car remains constant. The stress indicator χ_{ikt} is whether inventory has fallen more than 50% from its baseline level.

1.1.2 Producer Expectation Robustness [Work in Progress]

This normalized inventory is an imperfect proxy for a producer's expected months inventory of the marginal car. One reason is that a producer's optimal inventory level could change due to shifts in a producer's expectations of consumer demand. As such, I repeat the critical input specification using two different proxies for a producer's value of an extra car.

The first proxy, "Used Price (%)", is the average price of a used car for a given model as a fraction of a baseline used-car price. If a model's used car price is higher, presumably an additional new unit of that car model is more likely to sell. For example, used car prices spiked during COVID due to inventory shortages. One shortcoming of this proxy is that not all used cars are equally good substitute for their new counterpart. Since the used version of a less expensive make is probably a better substitute for the new version of the car, this should make the measure conservative. Another shortcoming is that manufacturers value of an extra car includes information about the future availability of the car which may not be priced in to used car markets.

For the second proxy, I similarly define "Subvention (%)" as the ratio of the subvention rate of a given car to a baseline subvention rate. Manufacturers lower subvention rates when inventory is low. For example, the subvention rate fell to nearly zero at the height of the shortages in 2021. As such, this metric holds information on how manufacturers view their inventory levels relative to anticipated demand. One shortcoming of this measure is that manufacturers all use subventions for different purposes. As such, baseline subvention rates vary significantly. Many baseline subvention rates are close to zero making any change in the subvention rate unobservable. Furthermore, the disparate reasons for subventions mean that manufacturers might also decrease subventions for reasons other than shortages of a given vehicle.

To do better than these two proxies for producers' value of an extra car, I would need to structurally estimate demand in the vein of Berry, Levinsohn, and Pakes (1995).

1.1.3 Missed Deliveries Specification [Work in Progress]

To level cars on the degree of supply chain disruptions, I will compare cars that are missing the same parts using data on missed deliveries. I will estimate equation 4 in its full form, where the scarcity

indicator, χ_{ikt} will be an indicator specifying whether a producer i missed a shipment of input k at time t. Liu, Smirnyagin and Tsyvinski (2024) propose a buyer-seller disruptions indicator. I will extend their methodology to make a supplier-producer-part disruption indicator. Focusing on fully missed shipments implies a disruption and not that producers are merely lowering sales expectations.

Individual parts will have a different input-producer-time fixed effect γ_{ikt}. As such, this specification measures producers' value of a given part (which can be substituted between a certain subset of a producers' cars). Although there are a huge number of parts, the part-specific fixed effects can be grouped across parts if the parts all fit into the same set of cars. Furthermore, the number of disruptions should be small enough that most of these fixed effects are ignored as the scarcity indicator will be turned off.

Since disruptions are at the part level, this specification removes the confounding factor that suppliers might strategically fortify supply chains for the most important parts which would likely support production of the more expensive cars. By honing in on the source of the disruption, this specification asks: if a shock affects two models equally, does the shock cause producers to reallocate parts to the more expensive of the two models?

1.1.4 Pecking Order Levels

Thus far the methodology has focused on the make level analysis; however I use each specification to test the four levels of reallocation. The four levels of reallocation correspond to partitioning the agents' set of projects at different granularities. In terms of the specifications, this means different ways of grouping observations and specifying inputs fixed effects. At the most aggregated level, the producer level, quantities are by model and input fixed effects vary by producer. At the plant level, quantities are by model and plant while input fixed effects vary by plant. At the platform level, quantities are also by plant and model but input fixed effects vary by platform. At the model level, prices and production vary by trim and input fixed effects are at the model level.

1.2 Data

I use monthly production and inventory data from Ward's Intelligence. Ward's provides U.S. production data for a given model and plant. Furthermore it includes U.S. inventories by model. Although the data begins in 1990, in order to focus on the COVID period I limit the sample period to start in 2018. The sample ends in June 2024. This gives enough observations before COVID

to establish a baseline period. I define this baseline period to be January 2018 to February 2020. Additionally, I remove models whose production averages less than 50 cars per months over the sample.

I use the SEC's ABS-EE data on auto loans to get an average invoice price on new cars for each generation of a given model.[1] I also use the ABS-EE dataset to get monthly data on the average used car price for a given model, and the cash subvention rate for new cars. After combining these two data sets, I am left with 125 models and 10 thousand observations.

1.2.1 Measuring Production and Producer Expectations [Work in Progress]

I measure production and producer expectations as a fraction of a baseline level. I define the baseline level to be the respective variable's average value during the baseline period (January 2018 to February 2020). This applies to the variables "Production (%)", "Inventory (%)", "Used Car (%)", and "Subvention (%)". In the case of production, the baseline represents the producers' target level of production. The assumption's interpretation is that producers designed plants to operate at an optimal capacity before COVID and were unable to adapt this optimal level of production during COVID. In the case of the variables measuring producer expectation, this normalization allows comparison across models. The assumption is that the optimal level of these variables does not change overtime.

1.2.2 Measuring Missed Deliveries [Work in Progress]

To measure missed deliveries for the missed deliveries specification, I will use S&P Panjiva Bill of Lading data, which includes shipment-level data on seaborn imports. The data includes the company name of the exporter and the importer, the port of lading and unlading, and details about the shipment's contents. S&P imputes 6-digit HS-Codes which categorize the contents of the shipment. Liu, Smirnyagin and Tsyvinski (2024) use the same data to make a buyer-seller disruption measure. I will extend this measure to be at the HS-Code level.

To compliment this disruptions measure, I will use part interchangeability data to determine which models share which parts. Since the S&P data classifies parts at the level of 6-digit HS-Codes, I will also classify interchangeability at this level of granularity. For each 6-digit HS-Code, I will partition cars based on whether they use the same part. Each subset within the partition

[1] Since Wards does not break down production by model generation, I infer model release dates from ABS-EE and change the model's price upon the release of a new generation.

of each 6-digit HS-Code corresponds to a part indexed by k. If a disruption happens to a given 6-digit HS-Code, then the stress indicator will activate for all parts within that 6-digit HS-Code. As a result, cars will be compared against other cars that use the same part within the disrupted 6-digit HS-Code.

1.3 Results [Work in Progress]

1.3.1 Results: Critical Input Specification

Table 1 displays the results from the critical input specification (Equation 1). Column (1) shows the results at the make level. The coefficient on inventory shows that greater inventory is correlated with greater production, meaning that this inventory control is serving its purpose of leveling cars on how badly they were hit by shortages. The coefficient on the interaction between price and inventory is -0.00606 (-16.06). Its interpretation is that the production of more expensive cars increases by more in response to decreases in inventory. The magnitude of the coefficient implies that if one car has a price of 10 thousand dollars more than another car and both cars' inventories collapse to zero, the more expensive car's production will fall by 5% less of its pre-COVID baseline production. In COVID, many cars' inventory did collapse to zero, so this 5% is significant in economic magnitude. In ongoing work, I will give a better interpretation of the magnitude of this coefficient by considering the counterfactual of a secondary market.

Column (2) shows the same result at the plant level. Many observations are dropped since many factories only produce one car, so the number of observations is actually lower than in the first column. Despite comparing production within factories as opposed to within makes, the coefficient on the interaction of production with inventory stays at a similar magnitude at -0.00406 (-5.25).This suggests much of the differential response of production to changes in inventory by car price can be explained by decision-making within factories.

1.3.2 Results: Producer Expectation Robustness [Work in Progress]

1.3.3 Results: Secondary Market Counterfactual [Work in Progress]

1.4 Figures and Tables [Work in Progress]

Figure 1: Inventory by Vehicle Price Bucket

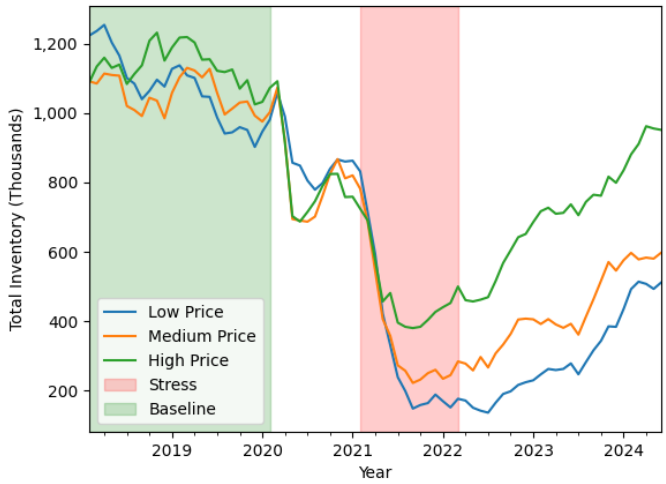


Figure 2: Average Price of Vehicle in Inventory

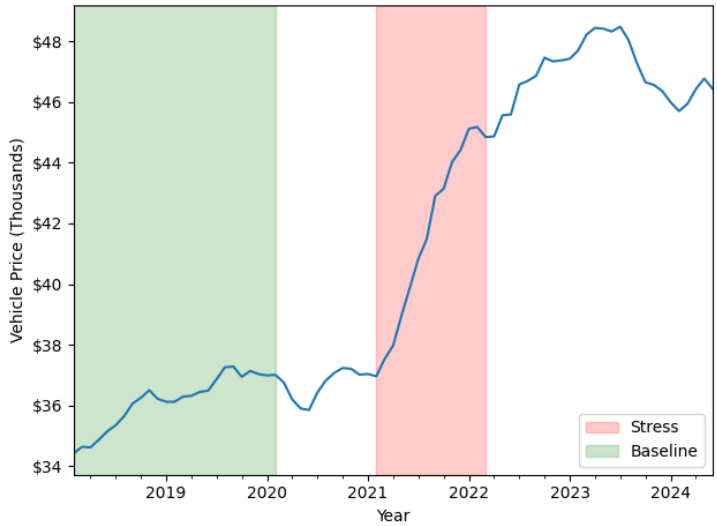
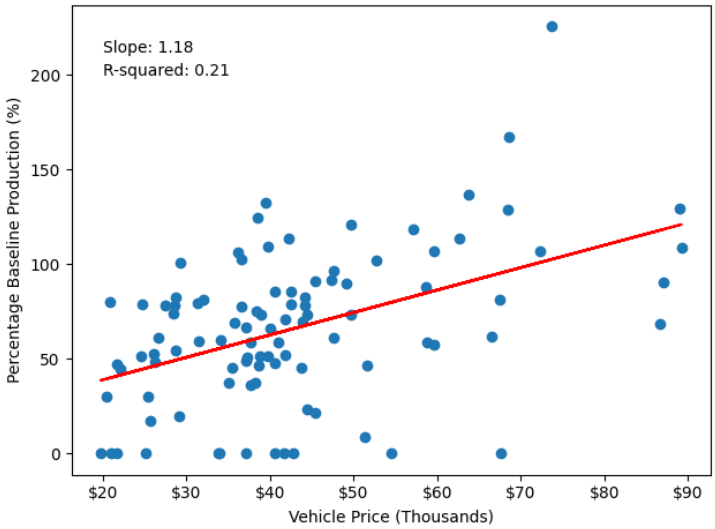


Figure 3: Percentage of Baseline Production by Vehicle Price



Cirtical Input Specification

	(1)	(2)
Dependent variable: Production (%)		
Level:	Manufacturer	Plant
Vehicle Price*Inventory (%)	-0.00606***	-0.00406***
	(-16.06)	(-5.250)
Inventory (%)	0.845***	0.689***
	(17.17)	(28.41)
Vehicle Price	2.23***	1.34***
	(10.17)	(7.130)
Constant	-60.2***	-16.3*
	(-4.733)	(-1.715)
Observations	7,296	6,337
R-Squared	0.629	0.636
Model FE	YES	YES
Manufacturer*Time FE	YES	
Plant*Time FE		YES

Robust t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix

A Model of Producer Inventory Surplus

I extend the model of excess capacity in Lewis-Hayre (2024) to answer: what is the optimal amount of inventory surplus for a producer? I adapt the model in Lewis-Hayre (2024) in the following ways. There are a countable number of periods $t \in \mathbb{N}$, where each time period behaves similarly to the base model. The state of the world $\omega_t \in \Omega$ is revealed at the start of each period. Agents are producers and projects are the goods that the producers produce. Instead of a single critical good, there are now a discrete set of inputs denoted K. Producers' value is profits denoted π_{it}. A producers' critical input target is its order for each input. Producers make orders for the next period. Denote θ_{ikt} to be producer i's order of input k scheduled to be delivered in period t but chosen in period $t-1$. Furthermore, producers must sell good j at p_{ijt} which is exogenous. Grossman, Helpman and Lhuillier (2023) focus on these pricing dynamics, whereas I focus on the impact of having common inputs. As such, I simplify away these pricing dynamics.[2]

Consumers are very simple. Each period a fixed number of consumers $s_{ijt}(\omega)$ are willing to pay p_{ijt} for good j. The true demand for each car is revealed to producers at the start of each period. Without changing the producer's problem, this share could depend on the price and availability of the goods produced by other consumers, but for now we ignore this possibility. As such, revenues are:

$$\mathcal{R}_{ijt}(\varsigma_{ijt}) = p_{ijt} \min\{\varsigma_{ijt}, s_{ijt}(\omega)\}$$

Where ς_{ijt} is the quantity of good j sold in period t.

Producer's operate a Leontief production function for each good. In particular:

$$y_{ijt} = \min_{k \in K_j} \alpha_{ijk} x_{ijkt}$$

Where y_{ijt} is the amount of good j produced at time t, $K_j \subseteq K$ are the inputs used to produce good j, x_{ijkt} is the quantity of input k required to produce one unit of good j, and α_{ijk} for $j \in K_j$ are the coefficients of the production function.

[2]For the empirical case of automobiles, static pricing also aligns with the stylized fact that MSRP does not much change over the course of the year. Even changes in loan interest rates are mostly explained by the yield curve. Although cash subventions fall close to zero in times of stress, they are small relative to the price of the car.

A producer faces costs from ordering inputs and running the production function. It also pays production adjustment costs. The production adjustment costs represent that it is difficult to change production plans. They take the form:

$$C(y_{ijt}) = \frac{\eta}{2}(\frac{y_{ijt} - \bar{y}_{ij}}{\bar{y}_{ij}} - \epsilon_{ijt})^2$$

where $\bar{y}_{ij}$ is an exogenously given optimal production level and ϵ_{ijt} is an i.i.d random shock to production costs. The interpretation is that producers prepare for a target level of production and deviations from this optimal level incur costs.[3] The total cost of producing good j is:

$$\mathcal{C}_{ijt}(\theta_{ikt}, y_{ijt}) = \sum_{k \in K_j} c_k \theta_{ikt} + \sum_{j \in J_i} c_{ij} y_{ijt} + \sum_{i \in I_j} C(y_{ijt})$$

Where c_k is the cost of ordering good k, and c_{ij} is the cost of running the production function for good j. As such producer i's profits from good j take the form:

$$\pi_{ijt} = \mathcal{R}_{ijt}(\varsigma_{ijt}) - \mathcal{C}_{ijt}(\theta_{ikt}, y_{ijt})$$

And total profits are:

$$\pi_{it} = \sum_{j \in J_i} \pi_{ijt}$$

The last addition is that producers can carry over inputs and final goods to future periods. In particular let ι_{ijt} denote producer i's inventory of good j. Let ι_{ijt} denote producer i's inventory of input k. Then we have the following inventory constraints:

$$\iota_{ikt+1} \leq \iota_{ikt} + q_{ikt}(\vec{\theta}_t, \omega_t) - x_{ikt}$$

$$\iota_{ijt+1} \leq \iota_{ijt} + y_{ijt} - \varsigma_{ijt}$$

Where $x_{ikt} = \sum_{j \in J_i} x_{ijkt}$. All the variables must be non-negative.

These extensions to the base model yield the following producer's problem. Producer i maximizes expected discounted profits:

[3] A generalization of the model would let producers dynamically choose the target level; however, for my empirical purposes, the observation period is short enough to assume that frictions to adjustment keep the target fixed.

$$
\begin{aligned}
\max_{\{\{\{\theta_{ikt+1}\}_{k\in K_j}\},\{y_{ijt},\varsigma_{ijt}\}_{i\in I_j}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \mathbb{E}_0[\delta^t \pi_{it}] \\
\text{s.t. } y_{ijt} = \min_{k\in K_j} \alpha_{ijk} x_{ijkt} \\
\iota_{ikt+1} \leq \iota_{ikt} + q_{ikt}(\vec{\theta}_t, \omega_t) - x_{ikt} \quad \forall k \in K_i \\
\iota_{ijt+1} \leq \iota_{ijt} + y_{ijt} - \varsigma_{ijt} \quad \forall j \in J_i \\
\iota_{ikt+1} \geq 0, x_{ikt} \geq 0, \theta_{ikt+1} \geq 0 \quad \forall k \in K_i \\
\iota_{ijt+1} \geq 0, y_{ijt} \geq 0, \varsigma_{ijt} \geq 0 \quad \forall j \in J_i
\end{aligned}
\tag{2}
$$

Where θ_{ik0} and ι_{ij0} are given $\forall k \in K_j$, and ι_{ij0} is given $\forall i \in I_j$. And δ is the discount factor.

A.1 Adjustment Cost Identification Strategy

To test the pecking order, I use the first order conditions for the production of a given good. The first order conditions reveal an identification strategy to measure the production adjustment costs associated with the pecking order. The identification strategy asks: how much extra do producers prioritize expensive cars when supply disruptions become more severe? I look to disprove the null hypothesis that production adjustment costs are prohibitively large.

The first order condition for y_{ijt} in the producer's problem is:

$$
0 = \xi_{ijt} + \mu_{ijt} - \sum_{k\in K_j} \xi_{ikt} - \sum_{k\in K} \lambda_{ik} - C'_{ij}(y_{ijt}) - c_j \tag{3}
$$

Where ξ_{ijt} is the value of extra inventory of good j today (or the Lagrange multiplier for the non-negativity constraint on ι_{ijt+1}, the inventory of good j); μ_{ijt} is the value of saving inventory of good j until tomorrow (or the Lagrange multiplier for the inventory constraint of good j); ξ_{ikt} is the value of extra inventory of input k today (or the Lagrange multiplier on the non-negativity constraint on ι_{ijkt+1}, the inventory of part k); and λ_{ijk} is the value of saving an extra unit of input k (or the Lagrange multiplier on λ_{ijk}, which equals δc_k).

What is the value of having an extra unit of car i? It is represented by the terms $\xi_{ijt} + \mu_{ijt}$. The terms' interpretation is that the value of an extra car is either the discounted value of a car tomorrow, or the revenues from selling the car.

In other words, if we denote $\upsilon_{ijt} = \mathbb{1}_{\iota_{ijt+1}>0}$, then we get that the value of an extra unit of car i is:

$$(1 - \upsilon_{ijt})p_{ijt} + \upsilon_{ijt}\delta\mathbb{E}_t[V'(\iota_{ijt+1})]$$

I approximate that:

$$p_{ijt}\mathbb{E}_t[\delta^{\tau_{ijt}}] = (1 - \upsilon_{ijt})p_{ijt} + \upsilon_{ijt}\delta\mathbb{E}_t[V'(\iota_{ijt+1})]$$

Where τ_{ijt} is the number of months until it takes to sell the marginal unit of good j produced in time t. In other words, the value of an extra car is its discounted expected revenue.

Denote $\sigma_{ijt} = \mathbb{E}_t[\delta^{\tau_{ijt}}]$ and $\hat{y}_{ijt} = \frac{y_{ijt}}{\bar{y}_{ij}}$. Then the first-order-conditions become:

$$\hat{y}_{ijt} = \frac{1}{\eta}(p_{ijt}\sigma_{ijt} - \sum_{k \in K_j} \xi_{ikt} - \sum_{k \in K_j} \lambda_{ik} - c_{ij}) - \bar{y}_{ij} + \epsilon_{ijt}$$

This gives the empirical specification:

$$\hat{y}_{ijt} = \beta_1 p_{ijt}\sigma_{ijt} + \sum_{k \in K_j} \gamma_{ikt}\chi_{ikt} + \sum_{k \in K_j} \gamma_k + \gamma_{ij} + \epsilon_{ijt} \tag{4}$$

Where $\chi_{ikt} = \mathbb{1}_{\iota_{ikt+1}=0}$ is an indicator specifying whether part k is scarce to producer i at time t, and ϵ_{ijt} is the error term. The coefficient of interest is β_1, and the null hypothesis is that $\beta_1 = 0$. The specification has model fixed effects, γ_{ij}, which represent the marginal cost of the car; input fixed effects, γ_k, which represent the cost of an input; and input-producer-time fixed effects, γ_{ikt}, which are only activated when a part is scare. The scarcity fixed effects represent the shadow cost (above the normal cost) of good j at time t to producer i. I refer to $\hat{y}_{ijt}$ as "Production (%)"

Since the expected months inventory of the marginal car, σ_{ijt}, and the scarcity of an input, χ_{ikt}, are not directly observable, I will approximate them. All information about demand enters through these two variables, so accurately measuring them could alleviate endogeneity concerns caused by producers' demand expectations.